THREE ESSAYS IN HEALTH ECONOMICS

by

Eugenia Amporfu
B.A., St. Francis Xavier University, 1994
M.A., University of Victoria, 1996

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In the
Department
of
Economics

SIMON FRASER UNIVERSITY

Fall 2004

APPROVAL

Name: **Eugenia Amporfu**

Degree: **Doctor of Philosophy**

Title of Thesis: **Three Essays in Health Economics**

Examining Committee:

Chair: **Dr. Ken Kasa**
Associate Professor , Department of Economics

____

Professor Gordon M. Myers
Senior Supervisor
Professor , Department of Economics

____

Professor Nicholas Schmitt
Supervisor
Professor , Department of Economics

____

Dr. Jane Friesen
Supervisor
Associate Professor, Department of Economics

____

Dr. Aslam Anis
External Examiner
Associate Professor, Department of Health and Epidemiology
University of British Columbia

____

Dr. Stephen Mongrain
Internal Examiner
Assistant Professor , Department of Economics

Date Defended/Approved: December 3rd, 2004

ABSTRACT

This dissertation consists of three independent essays addressing three separate health care policy issues.

Essay 1, "Incentive Effects of Government Mandated Cost-Shifting," shows how mandated cost shifting, because it does not require resources to pass through the hands of government, can be an optimal form of income redistribution in providing health care to the poor of society when government is sufficiently costly. Under this system, the government mandates the proper treatment of illness regardless of ability to pay and enforces that mandate with investigation. The paper shows that under costly information on illness the physician cheats by providing the wrong treatment when treating a rich patient who has low severity illness and a poor patient who has high severity illness. In response the government also investigates the treatment of such patients. The paper also shows the conditions under which mandated cost shifting is less wasteful and beneficial to patients.

Essay 2,"The Effects of the Relationship between Quantity and Quality of Care on Quality of Care," shows that the relationship between quality and quantity in the patient's utility as well as in the cost of care play an important role in determining the ability of a payment scheme to induce efficient quality and quantity of care. The payment schemes examined are fixed fee for service, prospective payment, and cost sharing. The paper shows that neither prospective payment nor fixed fee for service can be used to induce a first-best provision of quality and quantity. Cost sharing is the only scheme that can be used to induce the efficient supply of both quantity and quality.

Essay 3, "The Effect of Hospital Downsizing in British Columbia on the Quality of Care for Maternity Patients" uses maternity data from the Canadian province of British Columbia to estimate the effect of the reduction in hospital utilization rates and the transfer of care from hospitals to communities and to patients' homes on readmission rates. The results show that the policy reduced hospital length of stay and increased readmission rates for maternity patients.

DEDICATION

To Jesus Christ, my Love, who means more than this world to me. To my parents and Archbishop P. Sarpong, for continuing to believe in me.

ACKNOWLEDGEMENTS

The completion of this thesis is a further step in the deepening of my knowledge of God's faithfulness. Starting with the members of my committee, I would like to acknowledge those who were God's instruments in making this thesis a reality. Gordon Myers has been the stronghold of this thesis. His encouragement has been invaluable. He valued the idea in the first essay and encouraged me to pursue it. Working with him has really deepened my understanding of economic theory and modelling. This deepened understanding is also partly due to Nicholas Schmitt, whose financial support and dedication to the second chapter made an impression on me. Jane Friesen also helped to improve my empirical skills. This is very much appreciated.

I also would like to acknowledge the Knowledge Management and Technology Division of the British Columbia Ministry of Health for providing the data I used for the empirical chapter. I thank Reo Oddette for introducing me to SPSS without which I could not have written the last chapter of the thesis.

Finally, I would like to acknowledge the support of my friends. J. Atsu Amegashie stands out among them. Atsu was always willing to find time in his busy schedule to read and discuss my papers with me. J. Amoako-Tuffour stirred up my interest in Health Economics. Nicky Didicher and J. Boateng also found time to proofread the second and third chapters respectively. My community always prays for me. Nii Kwashie, Adolphine, Roko and Agatha never wearied of encouraging and praying for me. Worlanyo was always there to print drafts for me to read. Gift Dumedah was helpful in putting the thesis in the official format.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1: INCENTIVE EFFECTS OF GOVERNMENT MANDATED COST SHIFTING

1.1 Introduction

In every society there are poor individuals. A policy issue is the provision of the necessities of life to the poorest individuals in a society. One necessity is health care. If a society is not willing to see the poor die from treatable diseases then the provision of health care must involve some component of redistribution from the rich to the poor.[1] Even with improvements in heath-care technology, this is a problem, which may get worse. As health-care technology improves, the cost of treating some diseases may fall, but the set of treatable diseases will grow.

Different societies have used different re-distributive methods for making health care accessible to the poor. In Canada, for example, the health care act of 1945 was the adoption of public health or the no-fee public provision of health care to all residents of Canada. Public health care is the norm in much of continental Europe. The United Kingdom initially adopted a public health care system but then later allowed a parallel private for-fee system. The United States has never had a public health system but rather uses a private system with subsidized public hospitals and subsidized health care for the poor under Medicaid and the aged through Medicare.

A rather standard alternative economic proposal to the provision or subsidization of a service for redistributive reasons is to have the government use the tax system to directly redistribute income to allow people to cover their own expenses. This is not the practice in Canada and there are a number of arguments as to why. First, governments may want to ensure that the money is spent on health care rather than other goods (Evans, 1984). Second, given the reality of an incomplete set of lump-sum taxes, redistribution through the tax system will be distortionary and it is not theoretically obvious that public health provision is inferior to redistribution through the tax system. This has led to the application of the standard theoretical model for optimal taxation where the government maximizes social welfare subject to the government budget constraint. Blomqvist and Horn (1984) used this framework to determine optimal taxation and the circumstances under which it is optimal for government to transfer

[1] The poorest in any modern society cannot afford the actuarially fair premium to cover treatable diseases out of their own resources. But this paper will not be focused on issues of health insurance.

resources to sick people. Petretto (1999) used it to choose optimal tax, insurance and transfer in a health care system financed by a combination of social insurance, private insurance and patient co-payment (i.e. the patient is only partly insured and so directly pays part of the cost of services and the insurance pays the rest).

Under public health systems such as Canada's, the system is designed with the intention that access to life-saving treatment would not depend on ability to pay.[2] But in the United States, many developing countries (e.g. Ghana and Kenya), and in public health countries prior to the implementation of public health this was an obvious possibility. A question is; do people die from treatable disease due to an inability to pay in these societies? The answer is yes, but fewer than one might initially imagine. Take the case of Canada prior to the implementation of a public health system. Consider a community doctor faced by a patient with life-threatening disease that can be treated but only at a cost beyond the means of the patient. Were these patients allowed to die? Undoubtedly some were, but some doctors treated the patients, billed them, but then did not collect (Evans, 1984). At the time of implementation of public health in Canada some doctors' bills went uncollected. How would a doctor cover cost with revenue below costs for poor patients? The answer, of course, is with revenue above costs for rich patients, that is, with *cost shifting.* Cost shifting is also a practice in private American hospitals today.[3] It is a form of *private* redistribution from rich to poor orchestrated by doctors and hospitals. The obvious theoretical explanation for the phenomena lies in health-care practitioners caring about the well-being of their poor patients - they do not want to see them die from treatable diseases.

The systems discussed previously; public health care as in Canada, or subsidization of health care for the poor or aged as in the United States, or even direct redistribution through the tax system as in the optimal taxation literature, are public-sector intensive approaches when compared to cost-shifting. In particular under cost shifting the redistribution from rich to poor is orchestrated by the doctor using the price system so that no resources flow through the hands of the government. If the operation of government uses resources (e.g. the taxman's salary) or

---

[2] But it is possible to die from a treatable illness while waiting in a queue and the rich can avoid the queue by seeking treatment abroad.

[3] Numerous empirical studies have been done on cost shifting in hospitals (Dor and Farley, 1996; Zuckerman, 1987; Scheffler, Morrisey, 1995; Dranove, 1988; Morrisey and Sloan, 1989; Keating, 1984) and nursing homes (Little, 1992). While some did not find evidence of cost shifting (Showalter, 1997), others found evidence of cost shifting (Zuckerman, 1987, Dranove, 1988; Sloan and Becker 1984).

wastes resources (e.g. corruption) then this lack of public sector intensity can be a comparative benefit of the cost-shifting approach.[4]

The likely explanation for real-world cost-shifting is altruism for the poor by health-care practitioners. But if one assumes enough interdependent utility in a theoretical model then it is obvious that any redistribution problem can be alleviated. So in this paper I assume purely selfish agents so the government mandates the cost shifting by physicians. In developing countries such as Kenya and Ghana where patients make direct payments, poor patients are exempt from payment of certain services while the rich are expected to pay.

The purpose of this paper is to study mandated cost-shifting as a redistributive approach to the problem of providing adequate health care to the poor in the presence of costly government. The model extends Leger (2000) by introducing two categories of patients, rich and poor, combining the capitation in Leger with fee for services and replacing the insurance firm in Leger with a costly government that mandates the treatment of patients regardless of ability to pay. The patient's illness is either of high or low severity, which can be treated with high or low treatment respectively. Both patients pay the same fee for low treatment but the rich pay a higher fee for high treatment. Since the mandated cost shifting allows physicians to charge the rich a higher price than the poor it gives incentive to the selfish physician to cheat by providing inappropriate treatment to the patients. Government investigation is thus included to give the physician the incentive to treat the patient appropriately. The physician pays a fine if found guilty and the government gives any excess revenue after investigation to the poor as subsidy. In addition to doing comparative statics the paper also examines the inefficiency associated with cost shifting and how the resulting waste is affected by the parameters in the model.

As in Leger the equilibrium strategies can be solved in closed form. I show that mandated cost shifting benefits both rich and poor patients, relative to direct redistribution by tax, if the economy is able to adopt innovative technology that allows investigation at low cost or if the fine is sufficiently high because of the resulting low level of cheating. Patients are also well off under mandated cost shifting if the poor are high risk and the rich low risk patients. I also show that the cheating, the investigation, and fine collection under mandated cost shifting result in waste. Mandated cost shifting then functions efficiently in a rich economy that can afford innovative technology and where it is too costly to collect taxes. A poor economy is also better off under mandated cost shifting when it is too costly to collect taxes. If government is more likely to be

[4] See Burbidge and Myers (2004) on costly government.

costly in a developing economy than a developed economy then mandated cost shifting is more suitable for developing economy than a developed one.

There have been empirical studies that link cost shifting with moral hazard. Using data from the United States, Butler et al (1996) examines the reasons for increase in claims for soft tissue injuries in the 1980s. They found that 30% of the increase could be explained by moral hazard by physicians. Health Maintenance Organization (HMO) physicians are paid on the capitation basis but receive fee for service when the injury is work related. Physicians then can increase their income by classifying the injuries as work related, a process Butler et al. call cost shifting. Butler et al. (1997) examined how the rapid increase of HMOs affected the workers compensation cost per employee. They found that the expansion of HMOs caused a large increase in the workers compensation claim frequency. These studies differ from the current paper for the two reasons. First, moral hazard in these studies involves the misclassification of an illness as work related and they call this act cost shifting. Thus cost shifting itself is the moral hazard behaviour. Cost shifting in the current model is not cheating. Secondly, the moral hazard does not affect the kind of treatment that the patient receives and so the patient cannot be worse off by moral hazard. Cheating in the current paper involves treating a patient with the wrong treatment and is different from the cost shifting.

The paper is organized as follows. Section 2 outlines the model. Section 3 examines cost shifting and cheating under costless investigation. Section 4 is subdivided into two. Sections 4a examines the physician and the government's behaviours under private information on illness. Section 4b does comparative statics on cheating and government investigation. Finally Section 5 examines the inefficiency (the wastefulness of resources) related to cost shifting and Section 6 concludes.

1.2 The Model

This is a game with two players: a physician[5] and the government, and passive agents, patients, in a free entry and exit market.[6] The patient in Leger is replaced with two types of patients: rich and poor, indexed $j = R$ or P. Let r be the proportion of the rich among the patients and $(1 - r)$ be the proportion of the poor. As in Leger, a patient gets sick only once, drawing from two severities of illness: low severity (θ^L) and high severity (θ^H). The rich and the poor draw from

[5] The physician represents a health care provider and so can be a doctor, a hospital, a clinic, etc.

[6] The insurance company in Leger is replaced by government in this paper. Throughout the paper I assume that the patient is not fully informed about illness type.

the distribution $\theta = \{\theta^L, \theta^H\}$ with probabilities π_j and $(1- \pi_j)$ respectively. In accordance with Leger, the successful treatment of θ^H requires high treatment, e^H, and that of θ^L requires low treatment, e^L, with $e^L < e^H$. Income is observable to the doctor and the government. I examine two cases, costless investigation and costly investigation. Both the physician and the government can costlessly observe the type of illness under costless investigation but only the physician observes the type of illness costlessly with costly investigation. After observing the type of illness, the physician chooses treatment type. While this illness type is not costlelssly observable to the government, the treatment type is observable.

The government collects a lump-sum tax, γ_R, from rich patients to fund the investigation of the physician and for redistribution to the poor. It pays a lump-sum subsidy, γ_p, to the poor. A patient pays a fee, C^i ($i = H$ or L), for treatment and spends the rest of his income, $y_j \pm \gamma_j - C^i$,[7] on all other goods and services consumed. His health status, $\delta(\theta^i, e^i)$, is a function of illness and treatment. His utility function is:

$$U_j(y_j \pm \gamma_j - C^i, \delta(\theta^i, e^i))$$

$U_1 > 0$, $U_2 > 0$ and $U_{11} \le 0$, $U_{21} = U_{12} = 0$. The physician's utility is $V(C^i,e^i)$; $V_2 < 0$, $V_1 > 0$, $V_{11} < 0$, and $V(0,0) = 0$.[8]

In a perfectly competitive market with free entry and exit in which both patients and physicians are fully informed and patients pay for treatment, the physicians earn zero profit for each type of treatment and so $V(C^{L*}, e^L) = V(C^{H*},e^H) = 0$[9] and with $e^H > e^L$, it follows that $C^{H*} > C^{L*}$. However, without any redistribution I assume that the poor cannot afford C^{H*}. With $\gamma_p = 0$, I assume the poor cannot pay more than C^{L*} for health care.[10] A standard approach to such inequality would be for the government to redistribute income through the tax system, so that the poor can afford C^H or provide the health care directly by using tax revenue. However, such a solution is not necessarily optimal in this model because of the additional assumption of a costly government. The costliness of government can be due to high cost of the resources used, corruption, mismanagement, etc. Following Burbridge and Myers (2004), I assume that

---

[7] Where $y_j - \gamma_j = y_R - \gamma_R$ for $j = R$ and $y_j + \gamma_j = y_p + \gamma_p$.

[8] The assumption $U_{12} = 0$ is added to simplify the calculations. The removal of this assumption does not affect the results qualitatively.

[9] Note that $V(C^{H*}, e^H) = V(C^{L*}, e^L) = 0$ in a perfectly competitive equilibrium because if one is positive the physician increase her utility by providing the treatment type that provides positive utility. Besides with free entry and exit, other firms will enter and compete it away. Alternatively, if one is negative then physicians will exit the market.

[10] I assume that the poor cannot afford the actuarially fair insurance premium: $\pi_p C^L + (1-\pi_p)C^H$.

government is costly. The assumption is that for every dollar of revenue collected a given percentage, ω, is lost. In an extreme case in which $\omega = 1$ all revenue collected is lost through the costliness of government.

If the government uses the tax system then the government's budget is balanced if what is collected after waste is equal to what the poor receive, i.e., $r(1 - \omega)\gamma_R = (1- r)\ \gamma_p$ which implies that $\gamma_R = \dfrac{(1-r)\gamma_p}{r(1-\omega)}$. Thus γ_R approaches infinity as ω approaches one, for any finite, γ_p. The use of the tax system for redistribution in the presence of a costly of government then is not optimal.

Alternatively, I allow the government to redistribute income through the health care system by directly mandating the physician to give any patient the proper treatment and to charge the poor C^{L*} for either treatment and charge the rich C^{L*} for e^L.[11] The market then determine the high treatment fee, C^S paid by the rich, through free entry and exit of physicians or zero expected profit.

With both the rich and the poor paying C^{L*} for low treatment such that $V(C^{L*},e^L) = 0$, with $e^H > e^L$, and with the poor paying C^{L*} for e^H, it follows that $V(C^{L*}, e^H) < 0$, i.e., the physician makes a loss from treating a poor patient with e^H.[12] The physician can only earn a positive utility if $V(C^{S*}, e^H) > V(C^{H*}, e^H) = 0$ implying that $C^{S*} > C^{H*} > C^{L*}$.[13] However, this creates incentives for the physician to cheat when choosing treatment type. Thus, as in Leger, the physician may not always use the right type of treatment, but here the choice of treatment would depend on the patient's income type. I assume then that with probability $\alpha_j^{L/H}$ the physician chooses e^L to treat a patient of type j, given that the patient has drawn θ^H (cheats), and with probability $(1 - \alpha_j^{L/H})$ she treats the patient with e^H. With probability $\alpha_j^{L/L}$ the physician chooses e^L to treat a patient who has drawn θ^L and with a probability $(1-\alpha_j^{L/L})$ the physician treats the patient with e^H (cheats).

In order to protect the welfare of patients, I allow the government to investigate the physician.[14] The timing is as follows. A patient of an income type comes to a doctor and must be accepted.[15] The income type is observable to both the doctor and the government. The doctor costlessly observes the illness type. The doctor then chooses whether or not to cheat (α).

---

[11] We can imagine θ^L as the no illness state where e^L equals zero, so that $V(C^{L*},e^L) = 0$ implies $C^{L*} = 0$.
[12] The government could choose a different fee $C^H_p{}^*$ for the poor for high severity of illness, such that $V(C^H_p{}^*,e^H) < 0$ but that will not change the qualitative results.
[13] The focus of the paper is redistribution not insurance. The poor patient pays C^{L*} in either illness state so has no incentive to insure. The rich patient can be assumed risk neutral to avoid the insurance issue.
[14] In Leger, the insurance company undertakes investigation.
[15] This is a strong assumption but a structure where doctors are expected to do some treatment of the poor could be constructed.

Simultaneously, the government chooses whether or not to investigate conditional on the observed income type and treatment type e^H or e^L. Because the patient's illness type is not observable to the government without investigation, there is private information and the solution concept is Bayesian Nash equilibrium. Thus, I assume that the government investigates the physician with probability μ_j^L for a patient of income type j who has received e^L, and with probability $(1-\mu_j^L)$ it does not investigate. For e^H, it investigates with probability μ_j^H, and with $(1-\mu_j^H)$ it does not investigate.

If the government investigates it will be able to find out whether or not the physician used the right treatment but it incurs a cost, k. If the physician is found guilty of using the wrong effort she pays a fine, ϕ. As discussed in Becker (1968) and Shavell (1991), setting the fine sufficiently high would deter the physician from cheating. The problem with this is that of credibility (Andreoni, 1991). If the government sets the fine extremely high then the physician will not expect the government to implement it and so will cheat. Thus I only assume that the fine, ϕ, satisfies the conditions which require that the physician is better off not cheating than cheating and being caught with certainty:[16]

$$0 > V(C^{L*}, e^H) > V(C^{L*} - \phi, e^L) \quad \textbf{(1)}$$

$$0 = V(C^{L*}, e^L) > V(C^{S*} - \phi, e^H) \quad \textbf{(2)}$$

Equation **(1)** states that given the government investigates, the fine is high enough such that the physician is better off using e^H to treat a poor patient who has drawn θ^H, than treating him with e^L and being fined. Similarly, **(2)** states that, given the government investigates, the fine is high enough such that the physician is better off using e^L to treat a rich patient who has drawn θ^L, than treating him with e^H and being fined. I add the following assumptions:

Assumption 1: $U(y_j \pm \gamma_j - C^g, \delta(\theta^H, e^H)) > U(y_j \pm \gamma_j - C^g, \delta(\theta^H, e^L))$, $(g = S, L)$ a patient of type j with high severity of illness is better off when treated with e^H than when treated with e^L given the fees for type j.

Assumption 2: $U(y_j \pm \gamma_j - C^L, \delta(\theta^L, e^L)) > U(y_j \pm \gamma_j - C^g, \delta(\theta^L, e^H))$, a patient with low severity of illness is better off when treated with e^L than when treated with e^H given of fee paid.

[16] One advantage of cost shifting is that this usual credibility issue is less of a problem here because the physician who does not treat a severely ill patient, for financial reason, could expect a serious punishment.

Assumption 3: $k < (1 - \pi_p)(1 - \omega)\phi$ and $k < \pi_R(1 - \omega)\phi$. The expected net fine collected from investigation, given the physician is found guilty, exceeds the cost of investigation. The following summarizes the notation of conditional probabilities used:

Table 1.1: Notation of Conditional Probabilities

Symbol	Meaning	Meaning (in words)
$\beta_P^L(\alpha_P^{L/L},\alpha_P^{L/H})$	$\frac{\pi_P\alpha_P^{L/L}}{(1-\pi_P)\alpha_P^{L/H}+\pi_P\alpha_P^{L/L}}$	Probability that a poor patient has θ^L given that he has received e^L
$(1-\beta_P^L(\alpha_P^{L/L},\alpha_P^{L/H}))$	$\frac{(1-\pi_P)\alpha_P^{L/H}}{(1-\pi_P)\alpha_P^{L/H}+\pi_P\alpha_P^{L/L}}$	Probability that a poor patient has θ^H given that he has received e^L
$\beta_P^H(\alpha_P^{L/L},\alpha_P^{L/H})$	$\frac{(1-\pi_P)(1-\alpha_P^{L/H})}{(1-\pi_P)(1-\alpha_P^{L/H})+\pi_P(1-\alpha_P^{L/L})}$	Probability that a poor patient has θ^H given that he has received e^H
$(1-\beta_P^L(\alpha_P^{L/L},\alpha_P^{L/H}))$	$\frac{\pi_P(1-\alpha_P^{L/L})}{(1-\pi_P)(1-\alpha_P^{L/H})+\pi_P(1-\alpha_P^{L/L})}$	Probability that a poor patient has θ^L given that he has received e^H
$\beta_R^L(\alpha_R^{L/L},\alpha_R^{L/H})$	$\frac{\pi_R\alpha_R^{L/L}}{(1-\pi_R)\alpha_R^{L/H}+\pi_R\alpha_R^{L/L}}$	Probability that a rich patient has θ^L given that he has received e^L
$(1-\beta_R^L(\alpha_R^{L/L},\alpha_R^{L/H}))$	$\frac{(1-\pi_R)\alpha_R^{L/H}}{(1-\pi_R)\alpha_R^{L/H}+\pi_R\alpha_R^{L/L}}$	Probability that a rich patient has θ^H given that he has received e^L
$\beta_R^H(\alpha_R^{L/L},\alpha_R^{L/H})$	$\frac{(1-\pi_R)(1-\alpha_R^{L/H})}{(1-\pi_R)(1-\alpha_R^{L/H})+\pi_R(1-\alpha_R^{L/L})}$	Probability that a rich patient has θ^H given that he has received e^H
$(1-\beta_R^H(\alpha_R^{L/L},\alpha_R^{L/H}))$	$\frac{\pi_R(1-\alpha_R^{L/L})}{(1-\pi_R)(1-\alpha_R^{L/H})+\pi_R(1-\alpha_R^{L/L})}$	Probability that a rich patient has θ^L given that he has received e^H

I assume that the government's objective function, $EW(\alpha_j^{L/L}, \alpha_j^{L/H}, \gamma_j)$, is the sum of the expected utilities of the patients[17]:

[17] μ_j^i is not in the objective function of the government but is in the budget constraint below.

$$
\begin{aligned}
EW(\alpha_j^{L/L},\alpha_j^{L/H},\gamma_j) = (1-r)[(1-\beta_P^L)U_p(y_p+\gamma_p-C^L,\delta(\theta^H,e^L)) \\
+\beta_P^L U_p(y_p+\gamma_p-C^L,\delta(\theta^L,e^L)) \\
+\beta_P^H U_p(y_p+\gamma_p-C^L,\delta(\theta^H,e^H))+(1-\beta_P^H)U_p(y_p+\gamma_p-C^L,\delta(\theta^L,e^H))] \\
+r[\beta_R^L U_R(y_R-\gamma_R-C^L,\delta(\theta^L,e^L))+(1-\beta_R^L)U_R(y_R-\gamma_R-C^L,\delta(\theta^H,e^L)) \\
+\beta_R^H U_R(y_R-\gamma_R-C^S,\delta(\theta^H,e^H))+(1-\beta_R^H)U_R(y_R-\gamma_R-C^S,\delta(\theta^L,e^H))]
\end{aligned} \quad \textbf{(3)}
$$

The first two terms in the first square brackets represent a poor patient's expected utility when treated with e^L. The second two terms represent the patient's expected utility when treated with e^H. Similarly, the first two terms in the second square brackets represent the rich patient's expected utility when treated with e^L and the remaining two terms represent his expected utility when treated with e^H. The government's budget constraint is:

$$
\begin{aligned}
r(1-\omega)\gamma_R-(1-r)\gamma_P+(1-r)[\beta_P^L\mu_P^L(-k)+(1-\beta_P^L)\mu_P^L((1-\omega)\phi-k) \\
+\beta_P^H\mu_P^H(-k) \quad +(1-\beta_P^H)\mu_P^H((1-\omega)\phi-k)] \\
+r[\beta_R^H\mu_R^H(-k)+(1-\beta_R^H)\mu_R^H((1-\omega)\phi-k) \\
+\beta_R^L\mu_R^L(-k)+(1-\beta_R^L)\mu_R^L((1-\omega)\phi-k)]=0
\end{aligned} \quad \textbf{(4)}
$$

The first term in **(4)** represents the lump-sum tax collected from the rich after waste and the second term represents the lump-sum subsidy that is transferred to the poor. The first two terms in the first square bracket represent the expected net fine collected from the investigating the physician for treating a poor patient with e^L. The two remaining terms represent the expected fine collected from investigating the physician for treating the poor with e^H. The second square bracket represents expected fine collected from investigating the physician for the treatments provided to the rich with the first two terms for the high treatment and the last two for low treatment. The physician's expected utility if faced by a poor patient is:

$$
\begin{aligned}
EV(\alpha_j^{L/H},\alpha_j^{L/L},\mu_j^i) = (1-r)\{(1-\pi_P)[(1-\alpha_P^{L/H})V(C^L,e^H) \\
+\alpha_P^{L/H}(\mu_P^L V(C^L-\phi,e^L)+(1-\mu_P^L)V(C^L,e^L))] \\
+\pi_P[\alpha_P^{L/L}V(C^L,e^L)+(1-\alpha_P^{L/L})(\mu_P^H V(C^L-\phi,e^H) \\
+(1-\mu_P^H)V(C^L,e^H))]\}
\end{aligned} \quad \textbf{(5a)}
$$

The first square bracket in the first term represents her expected utility from treating a poor patient with e^L. The second square bracket represents her expected utility from treating the poor patient with e^H.

If faced by a rich patient the physician's expected utility is:

$$EV(\alpha_j^{L/H},\alpha_j^{L/L},\mu_j^i)=r\{\pi_R[(1-\alpha_R^{L/L})(\mu_R^H V(C^S-\phi,e^H)+(1-\mu_R^H)V(C^S,e^H))+\alpha_R^{L/L}V(C^L,e^L)] + (1-\pi_R)[\alpha_R^{L/H}(\mu_R^L V(C^L-\phi,e^L)+(1-\mu_R^L)V(C^L,e^L))+(1-\alpha_R^{L/H})V(C^S,e^H)]\} \quad \textbf{(5b)}$$

The first square bracket representing the doctor's expected utility from treating a rich patient with e^H and the second square bracket representing her expected utility from treating a rich patient with e^L.[18]

The government mandates the following behaviour from the physician: for all patients, treat θ^H with e^H and charge the poor C^{L*}; treat θ^L with e^L and charge C^{L*} whether the patient is rich or poor.

1.3 Costless Investigation Case

Assume the government does not incur any cost of investigation, $k = 0$.[19] The government chooses its strategies to maximize **(3)** subject to **(4)** with $k = 0$. Its strategies consist of choosing μ_j^i taking α_j^i and C^{S*}as given and choosing the optimal tax and subsidy. Notice that **(3)** is not a function of μ_j^i so the investigation is for revenue reasons only. The physician simultaneously chooses treatment type after observing the illness type, taking μ_j^i as given. The Bayesian Nash equilibrium strategies are derived in Appendix A.[20] The results are as follows:

$$1 > \mu_P^{H*} \geq 0, \qquad 1 > \mu_R^{L*} \geq 0$$

$$\mu_P^{L^*} \geq \overline{\mu}_P^{L^*} \equiv \frac{V(C^{L*},e^H)}{V(C^{L*}-\phi,e^L)} \qquad \textbf{(6)}$$

$$\mu_R^{H^*} \geq \overline{\mu}_R^{H^*} \equiv \frac{V(C^{S*},e^H)}{V(C^{S*},e^H)-V(C^{S*}-\phi,e^H)}$$

PROPOSITION 1: *Given that $k = 0$, the government investigates with probability less than one when it observes that a rich patient has received a low treatment or a poor patient has received a high treatment. However, the government may always investigate or investigates with a high probability when it observes that a rich patient has received a high treatment or a poor patient has received a low treatment.*

---

[18] The game is between a doctor with one patient and the government. But it can be more than one patient, N, by assuming enough linearity in the doctor's in the doctor's utility function or $NV(C, e) = V(NC^i, Ne^i)$. Example, $V(C, e) = C^i - e^i$.

[19] In other words, the government does not use any costly effort when investigating.

[20] After doing the $k > 0$ case, I will provide best response functions which will illustrate the logic of the results.

(Proof: Appendix A).

From **(1)** and **(2)**, $0 < \overline{\mu}_R^{L^*} < 1$, $0 < \overline{\mu}_R^{H^*} < 1$[21]. Intuitively, with $V(C^{S^*}, e^H) > V(C^{L^*}, e^L)$, the physician has incentive to cheat when treating a rich patient who has low severity of illness and a poor patient who has high severity of illness. Thus, the government will always investigate or investigate at a high probability because it incurs no cost in doing so and that it will collect revenue if the physician is found guilty. However, with $V(C^{L^*}, e^H) < V(C^{L^*}, e^L)$, the physician has no incentive to treat a rich patient who has high severity of illness with low treatment or a poor patient who has low severity of illness with high treatment. However, because investigation is costless, the government may investigate when it observes that a poor patient has been treated with high treatment and a rich patient with low treatment, i.e., $1 > \mu_p^{H^*} \geq 0$, $1 > \mu_R^{L^*} \geq 0$.

The physician chooses her strategies to maximize its expected utility taking the government's strategies as given. The strategies are derived in Appendix A and the results are as follows:

$$\alpha_p^{L/H^*} = (1 - \alpha_p^{L/L^*}) = (1 - \alpha_R^{L/L^*}) = \alpha_R^{L/H^*} = 0, \quad \textbf{(7)}$$

PROPOSITION 2: *Given $k = 0$, the physician plays the pure strategy of not cheating when treating patients.*

Proof: Appendix A.

Intuitively, with $V(C^{S^*}, e^H) > V(C^{L^*}, e^L)$, the physician has no incentive to treat a rich patient who has high severity of illness with low treatment; hence $\alpha_R^{L/H^*} = 0$. In the same way, with $V(C^{L^*}, e^H) < V(C^{L^*}, e^L)$, the physician has no incentive to treat a poor patient who has low severity of illness with high treatment; hence $\alpha_p^{L/L^*} = 1$. If the government observes the physician has used a wrong effort it fines the physician. The physician, knowing that the government can observe the type of illness costlessly, expects to pay a fine anytime she uses the wrong effort to treat a patient. By **(1)** and **(2)**, the physician is better off providing the right treatment than providing a wrong treatment and being fined. Thus, the physician will always provide the right effort.

With $k = 0$, the government does not need to collect revenue for investigation. Any revenue collected would be for redistribution of income. As shown in Appendix A, there is a critical $\omega = \omega^f$ where it is too costly for the government to redistribute income through tax system, i.e., $\gamma_R^* = 0$. Using the first-order conditions, γ_R^* is defined by:

[21] See The Appendix for proof.

$$2\frac{\partial U_P(y_P - C^{L*})}{\partial y_P} = \frac{1}{1-\omega^f}[\frac{\partial U_R(y_R - \gamma_R^* - C^{L*})}{\partial y_R} + \frac{\partial U_R(y_R - \gamma_R^* - C^{S*})}{\partial y_R}] \quad \textbf{(8)}$$

For redistribution, the costly government chooses the taxes such that the expected marginal utility of income for the poor equals the weighted (the weight = $(1/(1 - \omega))\in(1, \infty)$) expected marginal utility of income for the rich. The weight increases with the costliness of government. Intuitively, as ω increases from zero, the weight increases implying that the expected marginal utility of income for the poor is high and, with diminishing marginal utility, this means a decrease in the tax distributed to the poor. An increase in ω to ω^f also means that a greater percentage of revenue collected is lost and so only a small percentage can be transferred to the poor. The government then has to tax the rich more than what is required to make the poor afford high treatment. By making the rich patient pays a higher fee for high treatment the rich patient only pays what is required to allow the poor afford high treatment. Thus, as government becomes critically costly, it is optimal to distribute income through the health care system by forcing the rich to pay more for high severity of illness than the poor. Distributing income through taxes under such circumstance is too wasteful. Hence, $\gamma_j^* = 0$. See Appendix A for a proof.

The equilibrium is efficient or first best because there is no cheating and no costly investigation or taxation. Both patients receive the right treatment and the rich do not have to pay for investigation through the tax system. Assuming that each doctor's draw of patients out of the rich/poor patient distribution is that of the population (r, $1-r$) and that the doctor's utility function satisfies $V(C^i, e^i) = V(NC^i, Ne^i)$ and now substituting in **(7)**, C^{S*} can be determined by setting the physician's expected utility to zero (free entry and exit) or:

$$r(1-\pi_R)V(C^{S*}, e^H) + (1-r)(1-\pi_P)V(C^{L*}, e^H) = 0 \quad \textbf{(9)}$$

This determines $C^{S*}(C^L, \pi_R, \pi_P, r, e^H)$. These results are driven by the costless investigation on the type of illness. The assumption of $k > 0$ will give second-best results.

1.4 Costly Investigation Case

The government can find the illness type but only through investigation at cost $k > 0$. As before the government chooses μ_j^i, $\in (0,1)$, and γ_j, taking α_j^i as given, to maximize **(3)** subject to **(4)**. Similarly, the physician simultaneously chooses $\alpha_j^{L/L}$, $\in (0,1)$, and $\alpha_j^{L/H}$, $\in (0,1)$, taking μ_j^i as given, to maximize **(5)**. The government's equilibrium strategies are derived in the Appendix A and the results are as follows:

$$\mu_p^{H*} = \mu_R^{L*} = 0 \tag{10}$$

$$\mu_P^{L^*} = \overline{\mu}_P^{L^*}$$
$$\mu_R^{H^*} = \overline{\mu}_R^{H^*} \tag{11}$$

where $\overline{\mu}_P^{L^*}$ and $\overline{\mu}_R^{H^*}$ are defined in **(6)**.

PROPOSTION 3: *The government investigates the physician with probabilities that are strictly positive and less than one when it observes the physician has provided high treatment for a rich patient and low treatment for a poor patient. However the government does not investigate when it observes a poor patient is treated with high treatment and when a rich patient is treated with low treatment.*

(Proof: Appendix A).

From **(1)** and **(2)** μ_p^{L*} and μ_R^{H*} are positive and less than one and so are mixed strategies. The government plays the pure strategy of not investigating when it observes that a poor patient has received high treatment and a rich patient has received low treatment. The government however plays a mixed strategy when it observes that a poor patient has received low treatment and a rich patient has received high treatment.

Intuitively, $\mu_R^{L*} = 0$ because with $V(C^{S*},e^H) > V(C^{L*}, e^L)$, the physician has no incentive to treat a rich patient with e^L given that the patient has drawn θ^H. Similarly, $\mu_p^{H*} = 0$, i.e., the government does not investigate when it observes that a poor patient has been treated with e^H. As before, this is because with $V(C^{L*}, e^L) > V(C^{L*}, e^H)$, the physician has no incentive to use e^H to treat a poor patient given that the patient has drawn θ^L. However, with $V(C^{S*}, e^H) > V(C^{L*}, e^L)$, the physician has the incentive to treat a rich patient with e^H given that the patient has drawn θ^L. In the same way, with $V(C^{L*}, e^L) > V(C^{L*}, e^H)$ the physician has the incentive to treat a poor patient with e^L given that the patient has drawn θ^H.

The physician's strategies are:

$\alpha_p^{L/L*} = 1$, $\alpha_R^{L/H*} = 0$

$$\alpha_p^{L/H*} = \frac{\pi_p k}{(1-\pi_p)((1-\omega)\phi - k)}$$

$$(1-\alpha_R^{L/L*}) = \frac{(1-\pi_R)k}{\pi_R((1-\omega)\phi - k)} \quad \textbf{(12)}$$

PROPOSITION 4: *With probabilities that are positive but less than one, the physician provides high treatment to a rich patient given that he has drawn low severity of illness and provides low treatment to a poor patient given that he has drawn high severity of illness. However, the physician always provides low treatment to a poor patient given that the patient has drawn low severity of illness and provides high treatment to a rich patient given that the patient has drawn high severity of illness.*

(Proof: Appendix A).

Thus, $\alpha_p^{L/H*}$ and (1- $\alpha_R^{L/L*}$) are positive and by *Assumption 3* they are less than one. The physician again plays the pure strategy, of not cheating when treating a poor patient with low severity of illness and a rich patient with high severity of illness. The physician however plays mixed strategies when treating a poor patient with high severity of illness and a rich patient with low severity of illness. Note that as k approaches zero, the costless investigation results are obtained.

The propositions and results for $k = 0$ can be illustrated with best-response functions for the government and physician. Consider the case of a poor patient and use the obvious results that $\mu_p^{H*} \geq 0$, $\mu_R^{L*} \geq 0 = \alpha_R^{L/H*} = 0$, $\alpha_p^{L/L*} = 1$. Then it is straight forward to show that the governments objective is linear in μ_p^L and that the first order condition with respect to μ_p^L is linear and weakly increasing in $\alpha_p^{L/H}$. Further, it is straight forward to show that the doctor's objective is linear in $\alpha_p^{L/H}$ and the first order condition with respect to $\alpha_p^{L/H}$ is linear and weakly decreasing in μ_p^L. As a result the best response functions are step functions as in Figure 1 with a unique mixed strategy equilibrium at μ_p^{L*} and $\alpha_p^{L/H*}$. Finally, by letting k go to zero $\alpha_p^{L/H*}$goes to zero and the costless investigation results are as derived in **(6)**.

Figure 1.1: Costless investigation

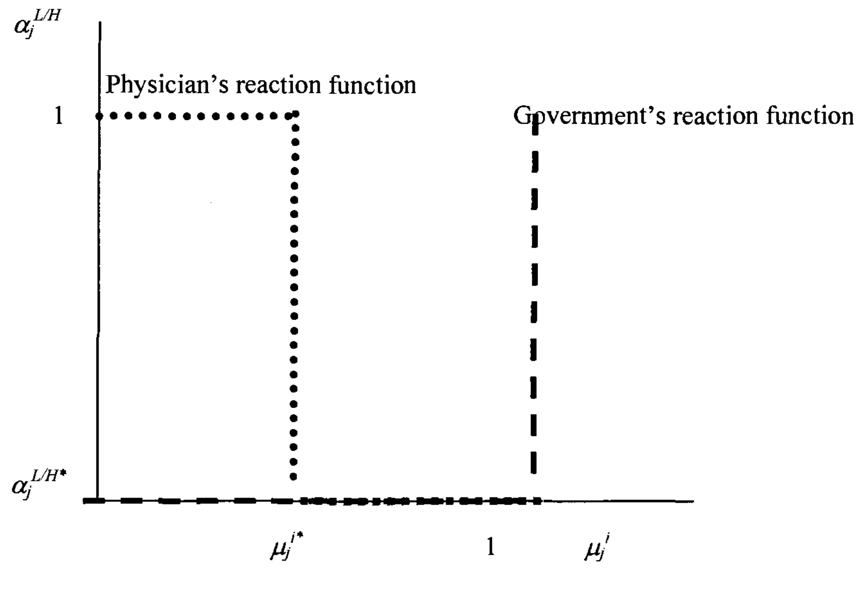
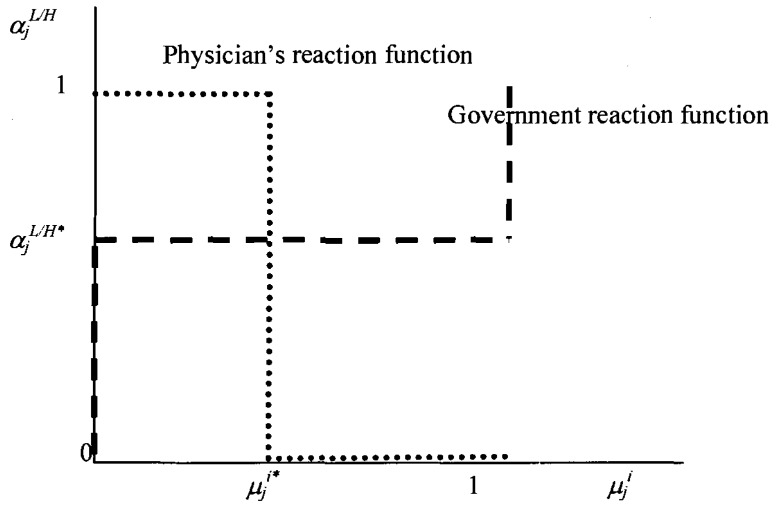


Figure 1.2: Costly investigation



With free entry and exit, the physician's expected utility goes to zero in equilibrium. Then following the same procedure as used for **(9)** but now using **(10)** and **(11)**, I get **(9)** again. Thus, $C^{S*}(C^L, \pi_R, \pi_p, r, e^H)$ from **(9)** is the same as that under costly investigation. This strong result comes from the mixed strategies. In equilibrium, the government chooses its strategies to make the physician indifferent between cheating and not cheating whether or not k is zero.

$$2\frac{\partial U_{y_p}(y_p - C^{L*})}{\partial y_p} = \frac{1}{1-\omega^A}\left[\frac{\partial U_{y_R}(y_R - \gamma_R^* - C^{L*})}{\partial y_R} + \frac{\partial U_{y_R}(y_R - \gamma_R^* - C^{S*})}{\partial y_R}\right] \quad \textbf{(13)}$$

Equation **(13)** defines the γ_R^*. Again there is a critical value of $\omega = \omega^A$ where $\gamma_R^* = 0$. From **(8)** and **(13)**, $\omega^f = \omega^A$. This means that the critical costliness of government at which a tax system is less efficient than cost shifting does not change with cheating. In the absence of cheating any waste from a tax system comes from the revenue collection but with cheating there is an additional waste from bad treatment and investigation. According to *Assumption 3*, given that the physician is found guilty, the fine collected after waste exceeds the cost of investigation. However, as shown in Appendix A, the expected net fine is zero because, whether the investigation is for the rich or the poor, the expected net fine collected when the physician is found guilty equals the expected cost of investigation that does not find the physician guilty. Thus, for the budget to balance, $\gamma_p^* = 0$.

The costly investigation equilibrium is Pareto inferior to the costless investigation equilibrium because those who get the right treatment are indifferent, but those who get the wrong treatment are worse off. The costly investigation case then is inefficient and the inefficiency consists of over-utilization of care for the rich and under-utilization of care for the poor.

The results here are similar to Leger in that the physician obtains the same expected utility as the first-best case but the patients are worse off. It, however, differs from Leger in that the capitation in Leger leads to under utilization of care while the combination of fee for service (for the rich) and capitation (for the poor) in this paper leads to a combination of over utilization and under utilization depending on the income of the patient.

Note that the difference between the costless and costly investigation cases in the current is costliness of investigation. The results so far imply that as investigation becomes more costly the equilibrium of mandated cost shifting becomes less efficient. As already explained, the free entry and exit assumption gives the physician no incentive to cheat when there is no income difference between patients. Thus direct redistribution of income through the tax system should achieve efficiency in terms of treatment. Such equilibrium is achievable under the mandated cost shifting if investigation is not costly. Efficiency then suggests that the government invests in innovative technologies to minimize the cost of investigation.

Before pursuing further with the efficiency characteristics of the model it is of interest to examine the conditions under which mandated cost shifting favours the poor and/or the rich. To do this I do comparative statics with respect to the parameters in the model.

1.5 Comparative Statics

Comparative statics allow me to examine what happens to the welfare of patients under mandated cost shifting, as a result of changes in the parameters. As already discussed, cheating makes the patients worse off and so any change in the parameters that reduce cheating represents an improvement in the patients' welfare.[22]

$$\frac{\partial \mu_p^{L*}}{\partial e^L} = -\frac{V(C^{L*}, e^H)V_2(C^{L*} - \phi, e^L)}{[V(C^{L*} - \phi, e^L)]^2} < 0$$
$$\frac{\partial \mu_p^{L*}}{\partial e^H} = \frac{V_2(C^{L*}, e^H)}{V(C^{L*} - \phi, e^L)} > 0 \qquad \textbf{(14)}$$

$$\frac{\partial \mu_R^{H*}}{\partial e^H} = \frac{\partial \mu_R^{H*}}{\partial C^{S*}}\frac{\partial C^{S*}}{\partial e^H} = -\frac{\partial \mu_R^{H*}}{\partial C^{S*}}\frac{[(1-r)(1-\pi_p)+r(1-\pi_R)]V_2}{r(1-\pi_R)V_1} > 0 \qquad \textbf{(15)}$$

where $\frac{\partial \mu_R^{H*}}{\partial C^{S*}}$ is positive.[23]

PROPOSITION 5: *An increase in the low (high) treatment effort decreases (increases) investigation. A change in the effort however has no direct effect on the physician's cheating.*

Intuitively, when the low treatment effort increases the gap between the low and high effort reduces and so the physician's incentive to cheat, by using low treatment for the poor with high severity of illness, falls. Hence investigation falls to make the physician indifferent between cheating and not cheating. However, when high treatment effort rises the government increases its investigation of low treatment because the physician has the incentive to avoid the risen high treatment effort when treating the poor. The government also increases its investigation of high treatment when the high treatment effort rises because the risen effort also increases the high treatment fee, which in turn increases the physician's expected utility from treating the rich with

[22] Note that the low treatment fee is not included in the parameters because it is not a parameter. When the treatment effort changes the low treatment fee adjust so that $V(C^{L*}, e^L) = 0$.

[23] $$\frac{\partial \mu_R^{H*}}{\partial C^{S*}} = \frac{V(C^S, e^H)V_1(C^S - \phi, e^H) - V(C^S - \phi, e^H)V_1(C^S, e^H)}{[V(C^S, e^H) - V(C^S - \phi, e^H)]^2} > 0$$

high treatment effort. Thus, the government increases investigation to make the physician indifferent between cheating and not cheating when treating a patient with high severity of illness.

All things being equal, treatment effort falls when there is improvement in technology. Treatment effort then captures treatment technology. The results then imply that under mandated cost shifting an improvement in treatment technology does not affect the welfare of patients in terms of a change in cheating but affects investigation such that cheating does not change. The improvement in the technology however benefits rich patients who receive high treatment because of the resulting reduction in the fee they pay.

I now examine the effect of the fine on the strategies.

$$\frac{\partial \mu_R^{H*}}{\partial \phi} = -\frac{V(C^{S*}, e^H)V_1(C^{S*} - \phi, e^H)}{[V(C^{S*}, e^H) - V(C^{S*} - \phi, e^H)]^2} < 0$$
$$\frac{\partial \mu_P^{L*}}{\partial \phi} = \frac{V(C^L, e^H)V_1(C^{L*} - \phi, e^L)}{[V(C^{L*} - \phi, e^L)]^2} < 0 \tag{16}$$

$$\frac{\partial \alpha_P^{L/H*}}{\partial \phi} = -\frac{\pi_P k(1-\omega)}{(1-\pi_P)((1-\omega)\phi - k)^2} < 0$$
$$\frac{\partial (1-\alpha_R^{L/L*})}{\partial \phi} = -\frac{(1-\pi_R)k(1-\omega)}{\pi_R((1-\omega)\phi - k)^2} < 0 \tag{17}$$

PROPOSITION 6: *An increase in the fine leads to a reduction of the probabilities of investigation and cheating.*

The fall in investigation comes from the expectation that increasing the fine discourages cheating and so reduces the need to investigate. Intuitively, an increase in the fine decreases the physician's payoff from cheating and so the government decreases investigation in order to move back to equilibrium, where she is indifferent between cheating and not cheating.

Cheating, when treating both the rich and the poor, decreases when the fine increases even though investigation falls. Intuitively, the increase in the fine increases the expected payoff from investigating and so cheating falls to make the government indifferent between investigating and not investigating.[24] Thus when the fine is set sufficiently high, cheating is low and mandated cost shifting will favour both the rich and the poor.

[24] If we assume that $V(C^g - \phi)$, where $g = L, S$; approaches negative infinite when ϕ approaches infinity then note that investigation and cheating both go to zero.

The cost of investigation has no effect on the government's strategies but affects the physician's strategies:

$$\begin{aligned} \frac{\partial \alpha_p^{L/H*}}{\partial k} &= \frac{\pi_p(1-\omega)\phi}{(1-\pi_p)((1-\omega)\phi-k)^2} > 0 \\ \frac{\partial (1-\alpha_R^{L/L*})}{\partial k} &= \frac{(1-\pi_R)(1-\omega)\phi}{\pi_R((1-\omega)\phi-k)^2} > 0 \end{aligned} \qquad \textbf{(18)}$$

PROPOSITION 7: *An increase in the cost of investigation increases cheating to make the government indifferent between investigating and not investigating and so the probability of investigation does not change.*

Such results are consistent with the mixed strategy equilibrium. When the cost of investigation increases the government's expected payoff from investigating falls and so the physician increases cheating to make the government indifferent between investigating and not investigating and so investigation does not change.

The adoption of innovative technology reduces the cost of investigation and so the results here imply that when the technology of investigation improves mandated cost shifting improves the welfare of all patients. The results from the costless investigation and costly investigation show that a reduction in the cost of investigation improves the efficiency of equilibrium and so the results here (on costliness of investigation) also imply that improvement in investigation technology makes mandated cost shifting redistribute income more efficiently.

The costliness of government also does not affect investigation but affects the physician's strategies:

$$\begin{aligned} \frac{\partial \alpha_p^{L/H*}}{\partial \omega} &= \frac{\pi_p k\phi}{(1-\pi_p)((1-\omega)\phi-k)^2} > 0 \\ \frac{\partial (1-\alpha_R^{L/L*})}{\partial \omega} &= \frac{(1-\pi_R)k\phi}{\pi_R((1-\omega)\phi-k)^2} > 0 \end{aligned} \qquad \textbf{(19)}$$

PROPOSITION 8: *An increase in the costliness of government increases cheating to make the government indifferent between investigating and not investigating and so the probability of investigation does not change.*

An increase in the costliness of government increases cheating. Intuitively, when government is costly a significant part of the revenue collected is lost and so reduces the government's expected payoff from investigating. Hence cheating increases to make the government indifferent between investigating and not investigating. Thus given that it is still too costly to collect taxes, an improvement in the technology of fine collection improves the benefit of mandated cost shifting to both the rich and the poor.

The probability of a poor patient having low severity of illness, however, affects both the government and the physician's strategies:

$$\frac{\partial \mu_R^{H^*}}{\partial \pi_p} = \frac{\partial \mu_R^{H^*}}{\partial C^{S^*}} \frac{\partial C^{S^*}}{\partial \pi_p} = \frac{\partial \mu_R^{H^*}}{\partial C^{S^*}} \frac{(1-r)V(C^{L^*}e^H)}{r(1-\pi_R)V_1(C^{S^*}, e^H)} < 0 \qquad \textbf{(20)}$$

$$\frac{\partial \alpha_p^{L/H^*}}{\partial \pi_p} = \frac{k}{((1-\omega)\phi - k)(1-\pi_p)^2} > 0$$
$$\frac{\partial (1-\alpha_R^{L/L^*})}{\partial \pi_p} = 0 \qquad \textbf{(21)}$$

PROPOSITION 9: *An increase in the probability that a poor patient has low severity of illness decreases the probability of investigating high treatment but has no effect on the probability of investigating low treatment. An increase in the probability that a poor patient has low severity of illness however increases cheating when treating the poor but has no effect on cheating when treating the rich.*

Intuitively, when the probability that a poor patient has low severity of illness is high then the government is not too 'surprised' when it observes low treatment to a poor patient and so may not investigate. The physician thus increases cheating when treating the poor. When the probability that a poor patient has low severity of illness increases the high treatment fee falls, which in turn decreases the probability of investigating high treatment of the rich.

The increase in cheating (when treating the poor) is consistent with the mixed strategy equilibrium. The physician chooses her strategies to make the government indifferent between investigating and not investigating. As the probability that a poor patient has low severity of illness increases, holding the physician's strategies constant, the conditional probability that a poor patient has high severity of illness given that he has received low treatment, $(1-\beta_p^L)$, decreases. This decreases the government's expected payoff from investigating. Thus, to make the government indifferent, cheating when treating the poor has to increase.

This is similar to Leger where an increase in the probability of getting low severity of illness increases the probability that the physician treats a patient with low treatment given that the patient has drawn high severity of illness. In Leger however this result is for all patients since he does not categorize patients into rich and poor. The result, however, is contrary to Pauly (1980), where the physician avoids high-risk patients under capitation. Since the physician in the current model cannot avoid a patient, Pauly's finding would predict that the physician would increase cheating when the probability that a poor patient has high severity of illness increases. Such a prediction does not occur in this model because of the presence of investigation.

Several studies have shown that rich people are more likely to be healthy than poor people (Fuchs, 1975; Ettner, 1996; McDonough et al., 1997). This implies a relatively high probability that a poor patient has high severity of illness, thus making mandated cost shifting more beneficial to the poor when the poor in the economy are high-risk patients because of the low cheating that accompanies it. This however comes at the cost of reduction in welfare for the rich because of the resulting increase in high treatment fee.

The probability that a rich patient has low severity of illness also affects both the government and the physician's strategies.

$$\frac{\partial \mu_R^{H*}}{\partial \pi_R} = \frac{\partial \mu_R^{H*}}{\partial C^{S*}} \frac{\partial C^{S*}}{\partial \pi_R} = \frac{\partial \mu_R^{H*}}{\partial C^{S*}} \frac{V(C^S, e^H)}{(1-\pi_R)V_1(C^S, e^H)} > 0 \qquad \textbf{(22)}$$

$$\frac{\partial \alpha_p^{L/H*}}{\partial \pi_R} = 0$$

$$\frac{\partial (1-\alpha_R^{L/L*})}{\partial \pi_R} = -\frac{k}{((1-\omega)\phi - k)\pi_R^{\;2}} < 0 \qquad \textbf{(23)}$$

PROPOSITION 10: *An increase in the probability that a rich patient has low severity of illness increases the probability of investigating high treatment but decreases the probability of cheating when treating a rich patient. There is however no effect on the probability of investigating low treatment or the probability of cheating when treating the poor.*

The increase in the probability of investigating high treatment is driven by the relationship between the high treatment fee and the probability that a rich patient has low severity of illness. The fall in the probability that a rich patient has low severity of illness increases the high treatment fee, which also increases in the probability of investigating high treatment. Intuitively, as the high treatment fee increases the physician's payoff from cheating increases and so the government increases its investigation.

The reduction in cheating is consistent with the mixed strategy equilibrium. In addition to increasing the probability that the government investigates high treatment, an increase in the probability that a rich patient has low severity of illness also increases the conditional probability that the rich patient has low severity of illness given that he has received high treatment, (1- β_R^H), holding cheating constant. This increases the government's payoff from investigating. To make the government indifferent between investigating and not investigating, the physician has to reduce cheating when treating the rich. Even though the increase in the high treatment fee makes the rich patient worse off, *Assumption 1* ensures that the patient is better off as a result of the reduction in cheating when the rich are likely to be less severely ill.

Finally, I examine the effect of the proportion of the rich in the population.

$$\frac{\partial \mu_R^{H^*}}{\partial r} = \frac{\partial \mu_R^{H^*}}{\partial C^{S^*}} \frac{\partial C^{S^*}}{\partial r} = \frac{\partial \mu_R^{H^*}}{\partial C^{S^*}} \frac{(1-\pi_P)V(C^{L^*},e^H) - (1-\pi_R)V(C^{S^*},e^H)}{r(1-\pi_R)V_1(C^{S^*},e^H)} < 0 \qquad \textbf{(24)}$$

PROPOSITION 11: *An increase in the proportion of the rich decreases the probability of investigating high treatment. A change in the proportion of the rich however has no effect on the physician's strategies.*

Intuitively, an increase in the proportion of the rich implies a decrease in the number of the poor to be subsidized and so the high treatment fee falls which in turn decreases the probability of investigating high treatment. The increase in investigation makes the physician indifferent between cheating and not cheating and so cheating does not change. This implies that under mandated cost shifting, as far as cheating is concerned, patients in a rich economy are not necessarily better off than those in a poor economy. The rich who receive high treatment are however better off because of the reduction of the fee.

To sum up, studies have shown that poor patients are more likely to be sick than rich patient, whether the economy is rich or poor. Mandated cost shifting then favours patients because cheating is less likely to occur when the poor are likely to be high-risk patients and the rich are likely to be low risk patients. This however implies that greater cost from high treatment is passed on to the rich in the form of high fee and so the rich are worse off. The rich then are especially worse off in a developing economy where the proportion of the rich is small.

The rich patient who receive high treatment are also likely to pay a lower fee in a rich economy than in a poor one because a rich economy is more likely to afford innovative technology that reduces high treatment effort and so lower the high treatment fee. However, as far as cheating is concerned patients in a rich economy are not better off than those in a poor economy. Holding everything else constant, patients receive the same treatment whether the economy is rich or poor.

In general, patients enjoy high level of welfare under mandated cost shifting if there is technological improvement that reduces cost, be it in investigation, treatment effort or fine collection. It is therefore important that when choosing mandated cost shifting the government invests in innovative technologies in these areas to ensure that patients enjoy high level of welfare. I now examine the efficiency of mandated cost shifting by examining the waste that results in equilibrium. But first I show how cost shifting here relates to the cost shifting literature.

1.6 Cost Shifting and Waste

Cost shifting refers to the practice by health care providers of raising prices paid by one group of patients in order to provide health care to another group at a lower price (Morrisey & Sloan, 1989). This is different from the textbook price discrimination where the profit-maximizing producer allocates the services between the categories of patients until the marginal revenues are equalized. Under profit maximization, as the fee paid by one category of patients decreases, providers reduce the cost of treating these patients by reducing the services they receive (Showalter, 1997) rather than increasing the price of the other category. Thus it is price discrimination that is consistent with profit maximization but not cost shifting.[25] Under cost shifting, when the price paid by one group of patients falls, the physician increases the price of the other group of patients to cover the cost of providing services to the group whose fee has fallen.

The positive probabilities of cheating in the current model are consistent with the cost shifting literature that uses profit maximization in that, by varying effort, the physician varies the quality of service according to ability to pay.[26] However, the positive but less than one probability implies that she does not always vary services according to ability to pay and so may provide high treatment to a poor patient even though the patient can only pay C^{L*}. With free entry and exit and the resulting zero-profit equilibrium, the physician cannot accept a fee below cost without practising cost shifting. Such behaviour is driven by the presence of investigation. As already explained, with $V(C^{L*}, e^{L}) > V(C^{L*}, e^{H})$, the physician has the incentive to treat a poor patient who has high severity of illness with low treatment. It is the threat of investigation that induces the physician to sometimes treat a poor patient who has high severity of illness with high treatment.

[25] Morriesey (1994) calls the textbook price discrimination static cost shifting and calls cost shifting, as defined in this paper, as dynamic cost shifting.

[26] Note that altruism is consistent with cost shifting. Altruism is not included in this case because this study focuses on the effect of investigation holding altruism constant. In the mixed strategy equilibrium the inclusion of altruism will affect the probability of investigation. Altruism is thus excluded to reduce the complexity that could result with it.

Cost shifting has to be practiced to cover the cost of treating the poor with high treatment and charging them low treatment fee. Thus when the low treatment fee, C^L, falls, the high treatment fee, C^{S*}, has to increase to cover the cost of providing high treatment to the poor. This is found by differentiating (9) with respect to C^L:

$$\frac{\partial C^{S*}}{\partial C^{L}} = -\frac{(1-r)(1-\pi_P)V_1(C^L, e^H)}{r(1-\pi_R)V_1(C^{S*}, e^H)} < 0 \quad \textbf{(25)}$$

Thus, by mandating physicians to treat all patients regardless of ability to pay, the government is inducing physicians to shift the cost of treating poor patients with high treatment on to the rich who receive high treatment. Consistent with the target income theory, cost shifting here is necessary to keep the physician's expected utility at zero in equilibrium. If C^{L*} falls, C^{S*} increases to keep expected utility at zero.

It has already been shown that there is inefficiency or waste associated with mandated cost shifting. As explained in Section 2, in the absence of cost shifting, there is no cheating but the rich receive the right treatment while all the poor receive low treatment since that is all they can afford. Cost shifting then allows the poor to receive high treatment but that comes at a cost of waste. The waste can be found by calculating the resources lost in the system.

To compute the waste, I assume that $V(C^{i*}, e^i)$ is linear, i.e., $V(C^{i*}, e^i) = C^{i*} - e^i$. Recall that $V(C^{H*}, e^H) = V(C^{L*}, e^L) = 0$ which implies under linearity that $C^{H*} = e^H$ and $C^{L*} = e^L$. I categorize waste according to source: treatment waste, investigation waste and revenue waste. Treatment waste results from the unnecessary provision of high treatment to the rich patients who have low severity of illness. In the absence of cost shifting, the rich patients receive the right treatment so no resources are wasted during treatment. Even though the poor patients receive only low treatment regardless of the type of illness, with $e^H > e^L$ the poor with high severity of illness receive less effort (not more) than they need and so there is no waste of resources[27]. Thus, treatment waste occurs from the cheating that results from cost shifting. From *Assumption 2*, such unnecessary treatment makes the rich patients worse off and it is wasteful because the extra resource, $e^H - e^L$, (with linearity this becomes $C^{H*} - C^{L*}$) could have been used for something else to improve welfare. The treatment waste then is:

$$r\pi_R(1-\alpha_R^{L/L*})(C^{H*} - C^{L*}) \quad \textbf{(26)}$$

[27] If anything, too little resources are used for treatment rather than too much.

Note that treatment waste is zero under the costless investigation case even though there is cost shifting. Thus, the cost shifting per se does not cause the waste. [28] The waste results when cost shifting is combined with costly investigation so that cheating occurs.

As the name implies, investigation waste comes from the costliness of investigation. Even though investigation occurs under the costless investigation case, no waste results because of the zero cost of investigation. Thus, resources are used for investigating the physician only in the costly investigation case. The cost of investigation then increases the total cost of treatment to society. The resulting waste from investigation is:

$$k(1-r)\mu_p^{L*}[(1-\pi_p)\alpha_p^{L/H*}+\pi_p]+r\mu_R^{H*}k[\pi_R(1-\alpha_R^{L/L*})+(1-\pi_R)] \qquad \textbf{(27)}$$

The first square bracket represents the probability of providing low treatment to a poor patient. The first term in the square bracket is the probability that the physician provides low treatment to a poor patient with high severity of illness and the second term represents the probability that the physician provides low treatment to a poor patient who has low severity of illness. The first term of **(27)** then is the expected cost of investigating the physician for treating a poor patient with low severity of illness. The second square bracket is probability of providing high treatment to the rich. The first term in the square bracket represents the probability of providing high treatment to the rich with low severity of illness and the second term is the probability of treating a rich patient with high severity of illness. The second term of **(27)** then is the expected cost of investigating the physician for treating the rich with high severity of illness.

Revenue waste comes from the costliness of government and occurs whenever the government collects revenue. Since $\gamma_R^* = 0$, the only revenue collected is the fine from the physician. The revenue waste is:

$$(1-r)\mu_p^{L*}\omega\phi(1-\beta_p^{L*})+r\mu_R^{H*}\omega\phi(1-\beta_R^{H*}) \qquad \textbf{(28)}$$

The first term represents the revenue waste from the fine collected when the physician is found guilty of cheating when treating the poor. The second term is that when the physician is found guilty of cheating when treating the rich. In the costless investigation equilibrium, $(1 - \beta_p^{L*}) = (1 - \beta_R^{H*}) = 0$ and so **(28)** is zero under costless investigation but positive under costly

[28] The rich pay $C^{S*} > C^{H*}$ for the high treatment and so C^{H*} pays the physician for the high treatment and $C^{S*} - C^{H*}$ represents a transfer from the rich to the physician, which eventually is transferred to the poor (zero expected profit for the physician) and so is not waste.

investigation.[29] With zero cheating under costless investigation, no fine is collected and so this waste is also zero. In an extreme case in which $\omega = 1$, the government does not investigate because it cannot collect the revenue to pay for the investigation. In such a case then investigation waste is zero.

The total waste, W, is the sum of **(26)**, **(27)** and **(28)**:

$$\begin{aligned} W = {} & r\pi_R(1-\alpha_R^{L/L*})(C^{H*}-C^{L*}) + k(1-r)\mu_p^{L*}[(1-\pi_p)\alpha_p^{L/H*}+\pi_p] \\ & + r\mu_R^{H*}k[\pi_R(1-\alpha_R^{L/L*})+(1-\pi_R)] \\ & + (1-r)\mu_p^{L*}\omega\phi(1-\beta_p^{L*}) + r\mu_R^{H*}\omega\phi(1-\beta_R^{H*}) \end{aligned} \quad \textbf{(29)}$$

Notice that $W = 0$ when $k = 0$.[30] To find out the extent to which mandated cost shifting is more wasteful under some economies than others, I do comparative statics on the total waste with respect to the parameters.

$$\begin{aligned} \frac{\partial W}{\partial r} = {} & \pi_R(1-\alpha_R^{L/L*})(C^{H*}-C^{L*}) - k\mu_p^{L*}[(1-\pi_p)\alpha_p^{L/H*}+\pi_p] \\ & - \left(\frac{(1-\pi_p)(C^{H*}-C^{L*})}{\phi(1-\pi_R)}\right)\{\omega\phi(1-\beta_R^{H*}) + k[(1-\pi_R)+\pi_R(1-\alpha_R^{L/L*})]\} \\ & - \phi\omega((1-\beta_p^{L*})\mu_p^{L*} < 0 \end{aligned} \quad \textbf{(30)}$$

Equation **(30)** shows that a change in the proportion of the rich has a negative effect on total waste when treatment waste is sufficiently small. The intuition here is that when investigation decreases significantly with an increase in the proportion of the rich patients then total waste decreases when the resulting decrease in investigation waste and revenue waste exceeds the increase in treatment waste. The investigation waste increases, even though cheating does not

---

[29] Recall that $(1-\beta_p^{L*}) = \dfrac{(1-\pi_p)\alpha_p^{L/H*}}{(1-\pi_p)\alpha_p^{L/H*}+\pi_p\alpha_p^{L/L*}}$ and $(1-\beta_R^{H*}) = \dfrac{\pi_R(1-\alpha_R^{L/L*})}{(1-\pi_R)(1-\alpha_R^{L/H*})+\pi_R(1-\alpha_R^{L/L*})}$ and so are equal to zero when $\alpha_p^{L/H*} = (1-\alpha_R^{L/L*}) = 0$ as is the case under the costless investigation equilibrium; but are positive when $\alpha_p^{L/H*} > 0, (1-\alpha_R^{L/L*}) > 0$ as under the costly investigation equilibrium.

[30] Note that if the government uses the tax system then the waste is $r\omega\gamma_R = \dfrac{r\omega(1-r)\gamma_p}{(1-\omega)}$ which approaches infinity when ω approaches one. Since the tax revenue is only required to cover the high treatment cost for the poor, $\gamma_p = C^{H*} - C^{L*}$.

increase, because of the rise in investigation. Thus, even though cheating is not affected by the richness of the economy, the results here show that mandated cost shifting is more efficient in a rich economy than a poor economy because the relatively low high treatment fee in a rich economy reduces investigation and, consequently, the resulting waste.

$$\frac{\partial W}{\partial k} = r[\pi_R(C^{H*} - C^{L*}) + \mu_R^{H*}k]\frac{\partial(1-\alpha_R^{L/L*})}{\partial k} + (1-r)\mu_p^{L*}[\pi_p + (1-\pi_p)(\alpha_p^{L/H*} + k\frac{\partial \alpha_p^{L/H*}}{\partial k})]$$
$$+ r\mu_R^{H*}[\pi_R(1-\alpha_R^{L/L*}) + (1-\pi_R)] + \omega\phi[\mu_p^{L*}(1-r)\frac{\partial(1-\beta_p^{L*})}{\partial k} + r\mu_R^{H*}\frac{\partial(1-\beta_R^{H*})}{\partial k}] > 0 \quad \textbf{(31)}$$

Equation **(31)** shows that an increase in the cost of investigation increases waste from all the three sources of waste. The treatment waste increases because cheating when treating the rich increases with the cost of investigation. The revenue waste also increases because the conditional probabilities of cheating both increase in the cost of investigation. Increase in the cost of investigation directly increases the investigation waste but also indirectly increases investigation waste through the resulting increase in cheating when treating the poor. In general an increase in the cost of investigation increases cheating and so waste increases as well. Thus, whether the economy is rich or poor, mandated cost shifting is less wasteful as the cost of investigation becomes cheaper.

$$\frac{\partial W}{\partial \omega} = r\pi_R[(C^{H*} - C^{L*}) + k\mu_R^{H*}]\frac{\partial(1-\alpha_R^{L/L*})}{\partial \omega} + (1-r)(1-\pi_p)\mu_p^{L*}k\frac{\partial \alpha_P^{L/H*}}{\partial \omega}$$
$$+ \phi[(1-r)\mu_p^{L*}(1-\beta_p^{L*}) + r\mu_R^{H*}(1-\beta_R^{H*})] \quad \textbf{(32)}$$
$$+ \omega\phi\left((1-r)\mu_p^{L*}\frac{\partial(1-\beta_P^{L*})}{\partial \omega} + r\mu_R^{H*}\frac{\partial(1-\beta_R^{H*})}{\partial \omega}\right) > 0$$

As shown in **(32)**, an increase in the costliness of government increases total waste as well. The increase in waste also comes from the effect of costliness of government on cheating. The increase in costliness of government indirectly increases treatment waste and investigation waste through the resulting increase in cheating. The revenue waste increases due to direct effect of the increased costliness of government and indirectly through the resulting increase in cheating. Even though mandated cost shifting is more suitable for an economy where the costliness of government is high, the more costly the government is the less efficient is mandated cost shifting.

$$\frac{\partial W}{\partial \phi} = r\pi_R[(C^{H*} - C^{L*}) + k\mu_R^{H*}]\frac{\partial(1-\alpha_R^{L/L*})}{\partial \phi} + (1-r)(1-\pi_p)\mu_p^{L*}k\frac{\partial \alpha_P^{L/H*}}{\partial \phi}$$

$$+ k(1-r)[(1-\pi_p)\alpha_P^{L/H*} + \pi_p]\frac{\partial \mu_P^{L*}}{\partial \phi} + kr[\pi_R(1-\alpha_R^{L/L*}) + (1-\pi_R)]\frac{\partial \mu_R^{H*}}{\partial \phi}$$

$$+ \omega\phi\left((1-r)\mu_P^{L*}\frac{\partial(1-\beta_P^{L*})}{\partial \phi} + r\mu_R^{H*}\frac{\partial(1-\beta_R^{H*})}{\partial \phi}\right) < 0 \qquad \textbf{(33)}$$

Equation **(33)** shows that an increase in the fine leads to a decrease in the level of waste. The treatment waste and revenue waste decrease indirectly through the resulting decrease in cheating. Investigation waste decrease partly because of a fall in investigation and partly because of the fall in cheating. Thus, if the fine is sufficiently high mandated cost shifting redistributes income with little waste.

$$\frac{\partial W}{\partial \pi_p} = -\frac{(1-r)^2\mu_p^{L*}k((1-\pi_p)\alpha_p^{H/L*} - 1)}{\pi_p}$$

$$+ r\left(\omega\phi(1-\beta_R^{L*}) + k\pi_R((1-\alpha_R^{L/L*}) + (1-\pi_R)\right)\frac{\partial \mu_R^{H*}}{\partial C^{S*}}\frac{\partial C^{S*}}{\partial \pi_p} < 0 \qquad \textbf{(34)}$$

Equation **(34)** shows that an increase in the probability that a poor patient has low severity of illness decreases waste. Only investigation waste and revenue wastes are affected. Treatment waste is not affected because treatment waste pertains to the unnecessary treatment of the rich. Investigation waste increases indirectly through the resulting increase in cheating when treating the poor. The investigation waste also decreases indirectly through the resulting fall in the high treatment fee. However, the effect of the high treatment fee is stronger than the effect of cheating and so investigation waste falls. Revenue waste decreases because of the fall in the probability of investigation. Less investigation means fewer fines are collected and the waste that results from fine collection falls. Thus even though mandated cost shifting benefits the poor patients if the poor are high risk patients the results here show that this benefit comes at the cost of high level of waste.

$$\frac{\partial W}{\partial \pi_R} = -\frac{rk}{((1-\omega)\phi - k)}[k\mu_R^{H*} + (C^{H*} - C^{L*})] - rk\mu_R^{H*}$$

$$+ r[k(\pi_R(1-\alpha_R^{L/L*}) + (1-\pi_R)) + \omega\phi(1-\beta_R^{H*})]\frac{\partial \mu_R^{H*}}{\partial C^{S*}}\frac{\partial C^{S*}}{\partial \pi_R} \gtrless 0 \qquad \textbf{(35)}$$

An increase in the probability that a rich patient has low severity of illness decreases treatment waste, increases revenue waste but may increase or decrease investigation waste. Investigation waste increases if the resulting increase in investigation is stronger than the resulting fall in cheating. Treatment waste decreases because of the resulting fall in cheating when treating the rich patients. The increase in revenue waste also comes from the resulting increase in investigation of high treatment. Thus an increase in the probability that a rich patient has low severity of illness causes total waste to increase if any increase in investigation waste and revenue waste exceeds the fall in treatment waste. Even though the rich are better off under mandated cost shifting when they are low risk patients, because of the fall in cheating, the improvement in welfare may come at a cost of inefficiency due to costliness of investigation and fine collection.

These results are summarized below:

PROPOSITION 11: *Cost shifting leads to waste when combined with costly investigation. The waste comes from treatment, investigation and revenue collection. An increase in the proportion of the rich, the fine, or the probability that a poor patient has low severity of illness reduces the level of waste.*[31] *However, the waste increases with an increase in the cost of investigation and the costliness of government and is ambiguously affected by an increase in the probability that a rich patient has low severity of illness.*

The important question here is, is the mandated cost shifting a better policy than direct redistribution of income through taxes. To answer this question, consider first the waste when income is redistributed directly through taxes: $r\omega\gamma_R = \dfrac{r\omega(1-r)(C^{H*}-C^{L*})}{(1-\omega)}$ which approaches infinity when the ω approaches one. Mandated cost shifting is more efficient than taxation if the waste under the tax system exceeds that under the mandated cost shifting. Similarly, taxation is more efficient than mandated cost shifting when the waste under mandated cost shifting exceeds that under taxation. Thus, there is a critical level of ω at which mandated cost shifting and taxation produce the same waste. This paper then has examined the case in which the costliness of government exceeds the critical level. Even though costliness of investigation makes the equilibrium of mandated cost shifting a second-best, direct redistribution through taxes results in a third-best equilibrium if government is critically costly. Moreover, the

[31] But by a footnote above, as the fine gets sufficiently large there would be no waste.

results in this section have shown that the waste under mandated cost shifting is low in a rich economy, when the fine is sufficiently high, or when innovative technologies are adopted.

1.7 Conclusion

When it is too costly for a government to collect revenue, mandated cost shifting as a means of providing health services to the poor can be less wasteful than using the tax system. This paper examined the case in which the government redistributes income through the health care system by mandating proper treatment regardless of ability to pay and investigates the physician's behaviour. The results show that the physician cheats when there is costly investigation. Cheating when treating the poor involves randomly using low treatment to treat a patient with high severity of illness while that of the rich involves randomly treating a rich patient who has low severity of illness with high treatment. The government investigation of these treatments is also random.

Even though mandated cost shifting benefits the poor if they are likely to be high-risk patients, because of the resulting low cheating, such benefit comes at a cost of increased waste from increased investigation and fine collection. It also benefits the rich if they are likely to be low risk patients because of the resulting low probability of cheating. Such benefit can also be wasteful because of the accompanying increase in investigation and fine collection.

Despite making patients in rich or poor countries better off, mandated cost shifting is less wasteful in a rich economy because of the low probability of investigation. Mandated cost shifting is even less wasteful in an economy where the adoption of innovative technology makes investigation and treatment cheap to undertake. Since a developed economy is more likely to afford such innovation, mandated cost shifting will be more efficient in a rich economy than in a poor one. However, if government is more likely to be critically costly in a developing economy than a developed economy then mandated cost shifting is suitable for a developing economy. The decision to adopt mandated cost shifting then is driven by the costliness of government and so if government is not costly mandated cost shifting is not an efficient choice even if the economy adopts well advanced investigation technology. Given that the government in a developed economy and a developing economy are equally costly, mandated cost shifting is more efficient in a developed economy than a developing economy. Obviously, it would be interesting to test for such results, but that is the subject of future research.

1.8 Reference List

Andreoni, J. (1991), "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?" *Rand Journal of Economics* 22, 385-95.

Becker, G.S. (1968), "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, 169-217.

Blomqvist, A. and H. Horn (1984), "Public Health Insurance and Optimal Income Taxation", *Journal of Public Economics*, 24, 353-71.

Burbridge, J.B. and G.M. Myers (2004), "Tariff Wars and Trade Deals with Costly Government," *Discussion Paper*, Department of Economics, Simon Fraser University.

Butler, R.J., D.L. Durbin, and N.M. Helvacian (1996), "Increasing Claims for Soft Tissue Injuries in Workers' Compensation: Cost Shifting and Moral Hazard," *Journal of Risk and Uncertainty*, 13(1), 73-87.

Butler, R.J, R.P. Hartwig and H. Gardner (1997), "HMOs, Moral Hazard and Cost Shifting in Workers' Compensation," *Journal of Health Economics*, 16, 191-206.

Dor, A. and D.E. Farley (1996), "Payment Source and the Cost of Hospital Care: Evidence from Multiproduct Cost Function with Multiple Payers," *Journal of Health Economics*, 15(1), 1-21.

Dranove, D. (1988), "Pricing by Non-Profit Institutions: Case of Hospital Cost Shifting," *Journal of Health Economics*, 7(1), 47-57.

Ettner, S. (1996), "New Evidence on the Relationship between Income and Health," *Journal of Health Economics,* 15, 67-85.

Evans, R.G. (1984), *Strained Mercy: The Economics of Canadian Health Care,* Toronto: Butterworths.

Fuchs, V.R. (1975), *Who Shall Live? Health Economics and Social Choice*, Expanded Edition, New York: Basic Books.

Gertler, P.J. (1989), "Subsidies, Quality, and the Regulation of Nursing Homes," *Journal of Public Economics,* 38, 33-52.

Keating, B (1984), "Cost Shifting: An Empirical Examination of Hospital Bureaucracy," *Applied Economics*, 16(2), 279-89.

Lee, J.A., H. Birnbaum, H., & C. Bishop (1983), "How Nursing Homes Behave: A Multi-Equation Model Nursing Home Behaviour," *Social Science Med.,* 17, 1897-1906.

Leger, P. T. (2000), "Quality Control Mechanisms Under Capitation Payment for Medical Services," *Canadian Journal of Economics,* 33(2), 565-86.

Little, J.S (1992), "Public-Private Cost shifts in Nursing Home Care," *New England Economic Review*, 0(7), 3-14.

McDonough, G. J., D. Duncan, Williams and J. House (1997), "Income Dynamics and Adult Mortality in the United States, 1972 through 1989," *American Journal of Public Health,* 87,(9) 1476-83.

Morrisey, M.A. and F.A. Sloan (1989), "Hospital Cost Shifting and the Medicare Prospective Payment System," *Working Paper*, University of Alabama at Birmingham.

Morrisey, M.A. (1995), "Movies and Myths: Hospital Cost Shifting," *Business Economics*, 30(2), 22-25.

Norton, E. (1992), "Incentive Regulation of nursing Homes," *Journal of Health Economics*, 11, 105-28

Pauly, M. (1980), *Doctors and their Workshops*, Chicago: U. of Chicago Press.

Petretto, A. (1999), "Optimal Social Health Insurance with Supplementary Private Insurance," *Journal of Health Economics*, 18, 727-45.

Shavell, S. (1991), "Specific versus General Enforcement of Law," *Journal of Political Economics*, 99, 1088-1109.

Showalter, M. H. (1985), " Physician's Cost Shifting Behaviour: Medicaid versus other Patinets," *Contemporary Economic Policy*, 15(2), 74-84.

Sloan, F. A. and E. R Becker (1984), "Hospital Ownership and Performance," *Economic Inquiry,* 23(1), 21-36.

Zuckerman, S. (1987), "Commercial Insurers and All-Payer Regulation, Evidence on Hospital's Responses to Financial Need," *Journal of Health Economics*, 6, 165-87.

1.9 Appendix

1. Proof of Propositions 1 - 4

Deriving the strategies require working out the government's problem, the physician's problem, and the characterizing an equilibrium. In subsection 1.1, I do the government's problem, in 1.2 I do the physician's problem and the in 1.3, I characterize an equilibrium.

1.1. The Governments Problem

The Government's objective is denoted as *EW* and is:

$$
\begin{aligned}
EW(\alpha_j^{L/L},\alpha_j^{L/H},\gamma_j) &= (1-r)[(1-\beta_P^L)U_p(y_p+\gamma_p-C^{L*},\delta(\theta^H,e^L)) \\
&+\beta_P^L U_p(y_p+\gamma_p-C^{L*},\delta(\theta^L,e^L)) \\
&+\beta_P^H U_p(y_p+\gamma_p-C^{L*},\delta(\theta^H,e^H))+(1-\beta_P^H)U_p(y_p+\gamma_p-C^{L*},\delta(\theta^L,e^H))] \\
&+r[\beta_R^L U_R(y_R-\gamma_R-C^{L*},\delta(\theta^L,e^L))+(1-\beta_R^L)U_R(y_R-\gamma_R-C^{L*},\delta(\theta^H,e^L)) \\
&+\beta_R^H U_R(y_R-\gamma_R-C^S,\delta(\theta^H,e^H))+(1-\beta_R^H)U_R(y_R-\gamma_R-C^S,\delta(\theta^L,e^H))]
\end{aligned}
$$

The Government's budget constraint is denoted BC and is:

$$
\begin{aligned}
&r(1-\omega)\gamma_R-(1-r)\gamma_P+(1-r)[\beta_P^L\mu_P^L(-k)+(1-\beta_P^L)\mu_P^L((1-\omega)\phi-k) \\
&\quad+\beta_P^H\mu_P^H(-k)\quad+(1-\beta_P^H)\mu_P^H((1-\omega)\phi-k)] \\
&\quad+r[\beta_R^H\mu_R^H(-k)+(1-\beta_R^H)\mu_R^H((1-\omega)\phi-k) \\
&\quad+\beta_R^L\mu_R^L(-k)+(1-\beta_R^L)\mu_R^L((1-\omega)\phi-k)]=0
\end{aligned}
$$

which could be further simplified

$$
\begin{aligned}
&r(1-\omega)\gamma_R-(1-r)\gamma_P+(1-r) \\
&[\mu_P^L\frac{(1-\pi_p)\alpha_p^{L/H}((1-\omega)\phi-k)-\pi_p\alpha_p^{L/L}k}{(1-\pi_p)\alpha_p^{L/H}+\pi_p\alpha_p^{L/L}} \\
&\quad+\mu_P^H\frac{\pi_P(1-\alpha_P^{L/L})((1-\omega)\phi-k)-(1-\pi_P)(1-\alpha_P^{L/H})k}{(1-\pi_P)(1-\alpha_P^{L/H})+\pi_P(1-\alpha_P^{L/L})}] \\
&\quad+r[\mu_R^H\frac{\pi_R(1-\alpha_R^{L/L})((1-\omega)\phi-k)-(1-\pi_R)(1-\alpha_R^{L/H})k}{(1-\pi_R)(1-\alpha_R^{L/H})+\pi_R(1-\alpha_R^{L/L})} \\
&\quad+\mu_R^L\frac{(1-\pi_R)\alpha_R^{L/H}((1-\omega)\phi-k)-\pi_R\alpha_R^{L/L}k}{(1-\pi_R)\alpha_R^{L/H}+\pi_R\alpha_R^{L/L}}]=0
\end{aligned}
$$

And the non-negativity constraints $\mu_p^L\geq 0$, $\mu_p^H\geq 0$, $\mu_R^L\geq 0$, $\mu_R^H\geq 0$, $\gamma_p\geq 0$, $\gamma_R\geq 0$

and inequality constraints:

$$1-\mu_p^L\geq 0 \qquad \text{(IE1)}$$

$$1-\mu_p^H\geq 0 \qquad \text{(IE2)}$$

$$1-\mu_R^L\geq 0 \qquad \text{(IE3)}$$

$$1-\mu_R^H\geq 0 \qquad \text{(IE4)}$$

The government's problem is to maximize the Lagrangian

$$\varphi = EW + \rho_0 BC + \rho_1 IE1 + \rho_2 IE2 + \rho_3 IE3 + \rho_4 IE4$$

with the choices $\{\mu_p^L, \mu_p^H, \mu_R^L, \mu_R^H, \gamma_p, \gamma_R\}$ taken as given $\{\alpha_p^{L/L}, \alpha_p^{L/H}, \alpha_R^{L/L}, \alpha_R^{L/H}\}$, C^{L*} and C^S. Note that EW is not a function of μ_j^L, μ_j^H

The partial derivatives:

$$\frac{\partial \varphi}{\partial \mu_p^L} = -\rho_1 + (1-r)\rho_0 \frac{(1-\pi_p)\alpha_p^{L/H}((1-\omega)\phi - k) - \pi_p \alpha_p^{L/L} k}{(1-\pi_R)\alpha_p^{L/H} + \pi_R \alpha_p^{L/L}}$$

$$\frac{\partial \varphi}{\partial \mu_p^H} = -\rho_2 + (1-r)\rho_0 \frac{\pi_P(1-\alpha_P^{L/L})((1-\omega)\phi - k) - (1-\pi_P)(1-\alpha_P^{L/H})k}{(1-\pi_P)(1-\alpha_P^{L/H}) + \pi_P(1-\alpha_P^{L/L})}$$

$$\frac{\partial \varphi}{\partial \mu_R^H} = -\rho_3 + r\rho_0 \frac{\pi_R(1-\alpha_R^{L/L})((1-\omega)\phi - k) - (1-\pi_R)(1-\alpha_R^{L/H})k}{(1-\pi_R)(1-\alpha_R^{L/H}) + \pi_R(1-\alpha_R^{L/L})}$$

$$\frac{\partial \varphi}{\partial \mu_R^L} = -\rho_4 + r\rho_0 \frac{(1-\pi_R)\alpha_R^{L/H}((1-\omega)\phi - k) - \pi_R \alpha_R^{L/L} k}{(1-\pi_R)\alpha_R^{L/H} + \pi_R \alpha_R^{L/L}}$$

$$\begin{aligned}\frac{\partial \varphi}{\partial \gamma_R} = &-r \frac{\pi_R \alpha_R^{L/L}}{(1-\pi_R)\alpha_R^{L/H} + \pi_R \alpha_R^{L/L}} U_1(y_R - \gamma_R - C^{L*}, \delta(\theta^L, e^L)) \\ &+ \frac{(1-\pi_R)\alpha_R^{L/H}}{(1-\pi_R)\alpha_R^{L/H} + \pi_R \alpha_R^{L/L}} U_1(y_R - \gamma_R - C^{L*}, \delta(\theta^H, e^L)) \\ &+ \frac{(1-\pi_R)(1-\alpha_R^{L/H})}{(1-\pi_R)(1-\alpha_R^{L/H}) + \pi_R(1-\alpha_R^{L/L})} U_1(y_R - \gamma_R - C^{S*}, \delta(\theta^H, e^H)) \\ &+ \frac{\pi_R(1-\alpha_R^{L/L})}{(1-\pi_R)(1-\alpha_R^{L/H}) + \pi_R(1-\alpha_R^{L/L})} U_1(y_R - \gamma_R - C^{S*}, \delta(\theta^L, e^H)) + r(1-\omega)\rho_0\end{aligned}$$

$$\begin{aligned}\frac{\partial \varphi}{\partial \gamma_p} = &(1-r)[\frac{(1-\pi_P)\alpha_P^{L/H}}{(1-\pi_p)\alpha_P^{L/H} + \pi_P \alpha_P^{L/L}} U_1(y_p + \gamma_p - C^{L*}, \delta(\theta^H, e^L)) \\ &+ \frac{\pi_P \alpha_P^{L/L}}{(1-\pi_P)\alpha_P^{L/H} + \pi_P \alpha_P^{L/L}} U_1(y_p + \gamma_p - C^{L*}, \delta(\theta^L, e^L)) \\ &+ \frac{(1-\pi_P)(1-\alpha_P^{L/H})}{(1-\pi_P)(1-\alpha_P^{L/H}) + \pi_P(1-\alpha_P^{L/L})} U_1(y_p + \gamma_p - C^{L*}, \delta(\theta^H, e^H)) \\ &+ \frac{\pi_P(1-\alpha_P^{L/L})}{(1-\pi_P)(1-\alpha_P^{L/H}) + \pi_P(1-\alpha_P^{L/L})} U_1(y_p + \gamma_p - C^{L*}, \delta(\theta^L, e^H))] - (1-r)\rho_0\end{aligned}$$

$$\frac{\partial \varphi}{\partial \rho_0} = BC$$

Derivatives with respect to the multipliers return the inequality constraints. Because the government can always increase the utility of individuals with excess revenue $\rho_0^* > 0$

The Kuhn--Tucker conditions are:

$$\frac{\partial \varphi}{\partial \mu_p^L} \leq 0, \quad \mu_p^L \geq 0 \quad and \quad \mu_p^L \frac{\partial \varphi}{\partial \mu_p^L} = 0 \tag{A-1}$$

$$\frac{\partial \varphi}{\partial \mu_p^H} \leq 0, \quad \mu_p^H \geq 0 \quad and \quad \mu_p^H \frac{\partial \varphi}{\partial \mu_p^H} = 0 \tag{A-2}$$

$$\frac{\partial \varphi}{\partial \mu_R^H} \leq 0, \quad \mu_R^H \geq 0 \quad and \quad \mu_p^H \frac{\partial \varphi}{\partial \mu_R^H} = 0 \tag{A-3}$$

$$\frac{\partial \varphi}{\partial \mu_R^L} \leq 0, \quad \mu_R^L \geq 0 \quad and \quad \mu_R^L \frac{\partial \varphi}{\partial \mu_R^L} = 0 \tag{A-4}$$

$$\frac{\partial \varphi}{\partial \rho_1} = 1 - \mu_p^L \geq 0, \quad \rho_1 \geq 0 \quad and \quad \rho_1 \frac{\partial \varphi}{\partial \rho_1} = 0 \tag{A-5}$$

$$\frac{\partial \varphi}{\partial \rho_2} = 1 - \mu_p^H \geq 0, \quad \rho_2 \geq 0 \quad and \quad \rho_2 \frac{\partial \varphi}{\partial \rho_2} = 0 \tag{A-6}$$

$$\frac{\partial \varphi}{\partial \rho_3} = 1 - \mu_R^H \geq 0, \quad \rho_3 \geq 0 \quad and \quad \rho_3 \frac{\partial \varphi}{\partial \rho_3} = 0 \tag{A-7}$$

$$\frac{\partial \varphi}{\partial \rho_4} = 1 - \mu_R^L \geq 0, \quad \rho_4 \geq 0 \quad and \quad \rho_4 \frac{\partial \varphi}{\partial \rho_4} = 0 \tag{A-8}$$

$$\frac{\partial \varphi}{\partial \rho_0} = BC \geq 0, \quad \rho_0 \geq 0 \quad and \quad \rho_0 \frac{\partial \varphi}{\partial \rho_0} = 0 \tag{A-9}$$

$$\frac{\partial \varphi}{\partial \gamma_R} \leq 0, \quad \gamma_R \geq 0 \quad and \quad \gamma_R \frac{\partial \varphi}{\partial \gamma_R} = 0 \tag{A-10}$$

$$\frac{\partial \varphi}{\partial \gamma_p} \leq 0, \quad \gamma_p \geq 0 \quad and \quad \gamma_p \frac{\partial \varphi}{\partial \gamma_p} = 0 \tag{A-11}$$

Using the derivatives with respect to ρ_0 return the budget constraint.

1.2 The Physician's Problem

The physician's problem is to choose her strategies to maximizes the sum of (5a) and (5b).

$$\begin{aligned}EV = &(1-r)\{(1-\pi_P)[(1-\alpha_P^{L/H})V(C^L, e^H)\\&+\alpha_P^{L/H}(\mu_P^L V(C^L-\phi, e^L)+(1-\mu_P^L)V(C^L, e^L))]\\&+\pi_P[\alpha_P^{L/L}V(C^L, e^L)+(1-\alpha_P^{L/L})(\mu_P^H V(C^L-\phi, e^H)+(1-\mu_P^H)V(C^L, e^H))]\}\\&+r\{\pi_R[(1-\alpha_R^{L/L})(\mu_R^H V(C^S-\phi, e^H)+(1-\mu_R^H)V(C^S, e^H))+\alpha_R^{L/L}V(C^L, e^L)]\\&+(1-\pi_R)[\alpha_R^{L/H}(\mu_R^L V(C^L-\phi, e^L)+(1-\mu_R^L)V(C^L, e^L))+(1-\alpha_R^{L/H})V(C^S, e^H)]\}\end{aligned}$$

The non-negativity constraints are $\alpha_p^{L/H} \geq 0, \alpha_p^{L/L} \geq 0$, $\alpha_R^{L/H} \geq 0$, $\alpha_R^{L/L} \geq 0$ and the inequality constraints are:

$$1-\alpha_p^{H/L} \geq 0 \quad \text{(IE5)}$$

$$1-\alpha_p^{L/L} \geq 0 \quad \text{(IE6)}$$

$$1-\alpha_R^{H/L} \geq 0 \quad \text{(IE7)}$$

$$1-\alpha_R^{L/L} \geq 0 \quad \text{(IE8)}$$

The Physician's problem is to maximize the function:

$$\xi = EV + \lambda_1 IE5 + \lambda_2 IE6 + \lambda_3 IE7 + \lambda_4 IE8$$

with the choices $\{\alpha_p^{L/L}, \alpha_p^{L/H}, \alpha_R^{L/L}, \alpha_R^{L/H}\}$ taken as given taken as given $\{\mu_p^L, \mu_p^H, \mu_R^L, \mu_R^H, \gamma_p, \gamma_R\}$and C^S.

The partial derivatives are:

$$\frac{\partial \xi}{\partial \alpha_P^{L/H}} = (1-r)(1-\pi_P)[-V(C^{L*}, e^H) + \mu_P^L V(C^{L*}-\phi, e^L) + (1-\mu_P^L)V(C^{L*}, e^L)] - \lambda_1$$

$$\frac{\partial \xi}{\partial \alpha_P^{L/L}} = (1-r)\pi_P[V(C^L, e^L) - \mu_P^H V(C^L-\phi, e^H) - (1-\mu_P^H)V(C^L, e^H))] - \lambda_2$$

$$\frac{\partial \xi}{\partial \alpha_R^{L/H}} = r(1-\pi_R)[(\mu_R^L V(C^L-\phi, e^L) + (1-\mu_R^L)V(C^L, e^L)) - V(C^S, e^H)] - \lambda_3$$

$$\frac{\partial \xi}{\partial \alpha_R^{L/L}} = r\pi_R[V(C^L, e^L) - \mu_R^H V(C^S - \phi, e^H) - (1-\mu_R^H)V(C^S, e^H))] - \lambda_4$$

The derivative with respect to the multipliers returns the inequality constraints

The Kuhn--Tucker conditions are:

$$\frac{\partial \xi}{\partial \alpha_p^{L/H}} \le 0, \quad \alpha_p^{L/H} \ge 0, \quad \frac{\partial \xi}{\partial \alpha_p^{L/H}} \alpha_p^{L/H} = 0 \tag{A-12}$$

$$\frac{\partial \xi}{\partial \alpha_p^{L/L}} \le 0, \quad \alpha_p^{L/L} \ge 0, \quad \frac{\partial \xi}{\partial \alpha_p^{L/L}} \alpha_p^{L/L} = 0 \tag{A-13}$$

$$\frac{\partial \xi}{\partial \alpha_R^{L/H}} \le 0, \quad \alpha_R^{L/H} \ge 0, \quad \frac{\partial \xi}{\partial \alpha_R^{L/H}} \alpha_R^{L/H} = 0 \tag{A-14}$$

$$\frac{\partial \xi}{\partial \alpha_R^{L/L}} \le 0, \quad \alpha_R^{L/L} \ge 0, \quad \frac{\partial \xi}{\partial \alpha_R^{L/L}} \alpha_R^{L/L} = 0 \tag{A-15}$$

$$\frac{\partial \xi}{\partial \lambda_1} = 1 - \alpha_p^{L/H} \ge 0, \quad \lambda_1 \ge 0, \quad \frac{\partial \xi}{\partial \lambda_1} \lambda_1 = 0 \tag{A-16}$$

$$\frac{\partial \xi}{\partial \lambda_2} = 1 - \alpha_p^{L/L} \ge 0, \quad \lambda_2 \ge 0, \quad \frac{\partial \xi}{\partial \lambda_2} \lambda_2 = 0 \tag{A-17}$$

$$\frac{\partial \xi}{\partial \lambda_3} = 1 - \alpha_R^{L/H} \ge 0, \quad \lambda_3 \ge 0, \quad \frac{\partial \xi}{\partial \lambda_3} \lambda_3 = 0 \tag{A-18}$$

$$\frac{\partial \xi}{\partial \lambda_4} = 1 - \alpha_R^{L/L} \ge 0, \quad \lambda_4 \ge 0, \quad \frac{\partial \xi}{\partial \lambda_4} \lambda_4 = 0 \tag{A-19}$$

1.3. Equilibrium

1.3.a. Proof of Propositions 1 and 2

The following results are true for $k = 0$

From $\frac{\partial \xi}{\partial \alpha_p^{L/L}} = \frac{\partial EV}{\partial \alpha_p^{L/L}} - \lambda_2$ and by the assumption that $V(C^{L*}, e^L) = 0$ and $e^L < e^H$ we know $\frac{\partial EV}{\partial \alpha_p^{L/L}} > 0$ regardless of the size of μ_p^H. Then from the first order condition in (A-13), λ_2^* > 0. Which from (A-17) implies $\alpha_p^{L/L*} = 1$. Hence:

$$\alpha_p^{L/L*} = 1$$

$\lambda_2^* > 0$. From $\frac{\partial \xi}{\partial \alpha_R^{L/H}}$ and by the assumption that $V(C^{L*}, e^L) = 0$, and $V(C^{S*}, e^H) > 0$ and with $\lambda_3 \geq 0$ then $\frac{\partial \xi}{\partial \alpha_R^{L/H}} < 0$ regardless of the size of μ_p^H. Then the last condition in (A-14) implies $\alpha_R^{L/H*} = 0$. This implies that $\frac{\partial \xi}{\partial \lambda_3} > 0$ and using (A-18):

$$\alpha_R^{L/H*} = 0$$

and $\lambda_3^* = 0$.

With $\alpha_p^{L/L*} = 1$ and $\frac{\partial \varphi}{\partial \mu_p^H} = -\rho_2 \leq 0$ so by (A-2) and (A-6)

$$1 > \mu_p^{H*} \geq 0$$

and $\rho_2 = 0$.

With $\alpha_R^{L/H*} = 0$ and $\frac{\partial \varphi}{\partial \mu_R^L} = -\rho_4 \leq 0$ so by (A-4) and (A-8)

$$1 > \mu_R^{L*} \geq 0$$

and $\rho_4 = 0$.

Let $\alpha_p^{L/H*} > 0$. First note that $\frac{\partial \varphi}{\partial \mu_p^L} = \frac{\partial BC}{\partial \mu_p^L} - \rho_1$ and recall that $\alpha_p^{L/L*} = 1$. Then $\frac{\partial BC}{\partial \mu_p^L} > 0$ so that $\rho_1 > 0$ for $\frac{\partial \varphi}{\partial \mu_p^L} \leq 0$ that is (A-1). Then from (A-5) $\frac{\partial \varphi}{\partial \rho_1} = 0$ or $\mu_p^L = 1$. Using $\lambda_1 \geq 0$ and $\frac{\partial \xi}{\partial \alpha_P^{L/H}} = (1-r)(1-\pi_P)[-V(C^{L*}, e^H) + \mu_P^L V(C^{L*} - \phi, e^L)] - \lambda_1$ we have $\frac{\partial \xi}{\partial \alpha_P^{L/H}} < 0$, by (1). Then $\alpha_p^{L/H*} > 0$ and $\frac{\partial \xi}{\partial \alpha_P^{L/H}} < 0$ are not consistent with the last condition in (A-12). And using (A-16),

$$\alpha_p^{L/H*} = 0$$

and $\lambda_1^* = 0$. Let $\mu_p^L = 0$, and using $\lambda_1^* = 0$ implies $\frac{\partial \xi}{\partial \alpha_p^{L/H}} > 0$ which is not consistent with (A-11). So $\mu_p^L = 0$ is not consistent with the Kuhn-Tucker conditions. Given these and $\frac{\partial \xi}{\partial \alpha_P^{L/H}} \leq 0$ in (A-12), then $\mu_P^{L^*} \geq \frac{V(C^{L^*}, e^H)}{V(C^{L^*} - \phi, e^L)}$. Then from $\mu_p^{L^*} > 0$, $\frac{\partial \varphi}{\partial \mu_P^L} = 0$ from (A-1). From (1), $\mu_p^{L^*} < 1$ and from (A-5) $\rho_1^* = 0$. Therefore

$$0 < \mu_p^{L^*} \leq 1$$

and $\rho_1^* = 0$. Let $0 \leq \alpha_R^{L/L} < 1$. Note that $\frac{\partial \varphi}{\partial \mu_R^H} = \frac{\partial BC}{\partial \mu_R^H} - \rho_3$. Then $\frac{\partial BC}{\partial \mu_R^H} > 0$ so $\rho_3 > 0$ for (A-3) to be satisfied and $\mu_R^H = 1$ from (A-7). Also with $0 \leq \alpha_R^{L/L} < 1$ using (A-19) $\frac{\partial \xi}{\partial \lambda_4} > 0$ so $\lambda_4 = 0$. However, given these and (2), $\frac{\partial \xi}{\partial \alpha_R^{L/L}} = r\pi_R[-V(C^S - \phi, e^H)] > 0$ and so violates (A-15) so $\leq \alpha_R^{L/L} < 1$ is not consistent with Kuhn-Tucker conditions. Hence by (A-19)

$(1 - \alpha_R^{L/L^*}) = 0$

$\lambda_4^* = 0$. Given these and $\frac{\partial \xi}{\partial \alpha_R^{L/L}} \leq 0$ in (A-15), so $\mu_R^{H^*} \geq \frac{V(C^{S^*}, e^H)}{V(C^{S^*}, e^H) - V(C^{S^*} - \phi, e^H)}$ and from (2) and (A-7):

$0 < \mu_R^{H^*} \leq 1$

and $\rho_3^* = 0$.

This proves equations (6) and (7).

1.3.b. Proof of Propositions 3 and 4

The proofs for $\mu_R^{L^*}$, $\mu_p^{H^*}$, α_R^{L/H^*} and α_R^{L/L^*} are the same as when $k = 0$ and so will not be repeated here. The following then are proof for equations (11) and (12) for α_R^{L/L^*} and α_p^{L/H^*}

Let $\alpha_p^{L/H} = 0$ then $\frac{\partial \varphi}{\partial \mu_p^L} < 0$ so from (A-1) $\mu_p^L = 0$ and from (A-16) $\frac{\partial \xi}{\partial \lambda_1} > 0$ so $\lambda_1 = 0$. Given these $\frac{\partial \xi}{\partial \alpha_p^{L/H}} > 0$ which violates (A-12), so $\alpha_p^{L/H} > 0$ in equilibrium.

Let $\alpha_p^{L/H} = 1$. First note that $\frac{\partial\varphi}{\partial\mu_p^L} = \frac{\partial BC}{\partial\mu_p^L} - \rho_1$ and recall that $\alpha_p^{L/L*} = 1$. Then using *Assumption 3,* $\frac{\partial BC}{\partial\mu_p^L} > 0$ so that $\rho_1 > 0$ for $\frac{\partial\varphi}{\partial\mu_p^L} \leq 0$, that is (A-1). Then from (A-5) $\frac{\partial\varphi}{\partial\rho_1} = 0$ or $\mu_p^L = 1$. Using $\lambda_1 \geq 0$ and $\frac{\partial\xi}{\partial\alpha_P^{L/H}} = (1-r)(1-\pi_P)[V(C^{L*}-\phi, e^L) - V(C^{L*}, e^H)] - \lambda_1$, we have $\frac{\partial\xi}{\partial\alpha_P^{L/H}} < 0$, by (1). Then $\alpha_p^{L/H} = 1$ and $\frac{\partial\xi}{\partial\alpha_P^{L/H}} < 0$ are not consistent with the last condition in (A-12). And combining (A-16):

$$\mathbf{0 < \alpha_p^{L/H*} < 1}$$

and $\lambda_1^* = 0$. Given these and $\frac{\partial\xi}{\partial\alpha_P^{L/H}} = 0$ in (A-12), then $\mu_p^{L*} = \frac{V(C^{L*}, e^H)}{V(C^{L*}-\phi, e^L)}$. Using (1) and (A-5):

$$\mathbf{0 < \mu_p^{L*} < 1}$$

$\rho_1^* = 0$. Then from $\mu_p^{L*} > 0$ $\frac{\partial\varphi}{\partial\mu_p^L} = 0$ and using $\alpha_p^{L/L*} = 1$, $\alpha_P^{L/H*} = \frac{\pi_P k}{(1-\pi_P)((1-\omega)\phi - k)}$.

Let $\alpha_R^{L/L} = 0$ then $\frac{\partial BC}{\partial\mu_R^H} > 0$ so $\rho_3 > 0$ for (A-3) to be satisfied and $\mu_R^H = 1$ for (A-7). Also with $\alpha_R^{L/L} = 0$ using (A-19) $\frac{\partial\xi}{\partial\lambda_4} > 0$ so $\lambda_4 = 0$. However, given these and (2) $\frac{\partial\xi}{\partial\alpha_R^{L/L}} = r\pi_R[V(C^{L*}, e^L) - V(C^S - \phi, e^H)] > 0$, and so violates (A-15) so $\alpha_R^{L/L} = 0$ is not consistent with Kuhn-Tucker conditions.

Let $\alpha_R^{L/L} = 1$ then $\frac{\partial\varphi}{\partial\mu_R^H} < 0$ and to satisfy (A-3), $\mu_R^H = 0$. Given these, $\frac{\partial\xi}{\partial\alpha_R^{L/L}} = -r\pi_R V(C^S, e^H) - \lambda_4 < 0$, so $\alpha_R^{L/L} = 1$ violates (A-15) and so is not consistent with the Kuhn-Tucker conditions. Hence (A-19).

$$\mathbf{0 < (1 - \alpha_R^{L/L*}) < 1}$$

and $\lambda_4^* = 0$. Given these and $\frac{\partial \xi}{\partial \alpha_R^{L/L}} = 0$ in (A-15) $\mu_R^{H^*} = \frac{V(C^{S*}, e^H)}{V(C^{S*}, e^H) - V(C^{S*} - \phi, e^H)}$

Using (2) and (A-7):

$$\mathbf{0 < \mu_R^{H*} < 1}$$

and $\rho_3^* = 0$. With $\mu_R^{H*} > 0$ then $\frac{\partial \varphi}{\partial \mu_R^H} = 0$ from (A-3). From (A-3) and using $\alpha_R^{L/H*} = 0$ then:

$$(1 - \alpha_R^{L/L*}) = \frac{(1 - \pi_R)k}{\pi_R((1-\omega)\phi - k)}.$$

This proofs equations 11 and 12.

1.3.c. Proof of γ_j^*

Using the assumption that $U_{21} = U_{12} = 0$, $\frac{\partial \varphi}{\partial \gamma_p} = 2U_1(y_p + \gamma_p - C^{L*}) - \rho_0 \leq 0$. The assumption that $U_1 > 0$ implies $\rho_0^* > 0$ for (A-11) to be satisfied. Then from (A-9), $\frac{\partial \varphi}{\partial \rho_0} = BC = 0$.

Given $U_1 > 0$, $\rho_0^* > 0$ and using the assumption that $U_{21} = U_{12} = 0$, define ω^A $\frac{\partial \varphi}{\partial \gamma_R} = -[U_1(y_R - \gamma_R - C^{L*}) + U_1(y_R - \gamma_R - C^{S*})] + (1 - \omega^A)\rho_0 = 0$ at $\gamma_R = 0$ so when $\omega \geq \omega^A$

$$\boldsymbol{\gamma_R^* = 0}$$

Hence $\rho_0^* = \frac{U_1(y_R - \gamma_R - C^{L*}) + U_1(y_R - \gamma_R - C^{S*})}{(1 - \omega^A)}$

With $\alpha_R^{L/H*} \geq 0$, $(1 - \alpha_R^{L/L*}) \geq 0$, $\alpha_p^{L/L*} = 1$, $\alpha_R^{L/H*}$, $\mu_p^{L*} > 0$, $\mu_R^{H*} > 0$ $\mu_p^{H*} = 0$, $\mu_R^{L*} = 0$ and $\omega \geq \omega^A$ so $\gamma_R^* = 0$, BC becomes:

$$\begin{aligned} &-(1-r)\gamma_P \\ &+(1-r)\mu_P^L \frac{(1-\pi_R)\alpha_R^{L/H}((1-\omega)\phi - k) - \pi_R\alpha_R^{L/L}k}{(1-\pi_R)\alpha_R^{L/H} + \pi_R\alpha_R^{L/L}} \\ &+r\mu_R^H \frac{\pi_R(1-\alpha_R^{L/L})((1-\omega)\phi - k) - (1-\pi_R)(1-\alpha_R^{L/H})k}{(1-\pi_R)(1-\alpha_R^{L/H}) + \pi_R(1-\alpha_R^{L/L})} = 0 \end{aligned}$$

where each of the last two lines is zero and so

$$\gamma_p^* = 0$$

Given $\gamma_p^* = 0$ and using the assumption that $U_{21} = U_{12} = 0$, and substituting ρ_0^* into $\frac{\partial \varphi}{\partial \gamma_p}$

gives: $$\frac{\partial \varphi}{\partial \gamma_p} = 2U_1(y_p + \gamma_p^* - C^{L*}) - \frac{U_1(y_R - \gamma_R^* - C^{L*}) + U_1(y_R - \gamma_R^* - C^{S*})}{(1-\omega^A)} \leq 0$$

So $2U_1(y_p + \gamma_p^* - C^{L*}) = \frac{U_1(y_R - \gamma_R^* - C^{L*}) + U_1(y_R - \gamma_R^* - C^{S*})}{(1-\omega^A)}$ is consistent with the Kuhn Tucker conditions in (A-11).

This proves equations (8) and (13).

CHAPTER 2: THE EFFECT OF THE RELATIONSHIP BETWEEN QUALITY AND QUANTITY OF CARE ON QUALITY OF CARE SUPPLY

2.1 Introduction

One of the major concerns of reimbursers of health care providers is adopting the payment scheme that induces providers to provide quality of care at a level desirable to the reimburser. What makes this difficult is that quality of care is not contractible, and so insurance reimbursement is based on quantity of care such as number of visits or length of stay in the hospital. The provider is thus relied upon to invest and supply the (costly) quality effort required to improve the patient's health during a visit. To put it in economic jargon, there is a missing market for insurance and payment policies that is based on quality effort (Ma and McGuire, 1997). Since payment schemes are based on quantity, the relationship between quantity and quality of care is crucial in inducing the desired level of quality.

The purpose of this paper is to examine the role of the relationship between quantity and quality in the patient's utility and in the cost of production in determining the level of quality under three different payment schemes: fixed fee for service, prospective payment, and cost sharing. The effect of the relationship between quality and quantity on cost, cost-cross effect, refers to the change in the marginal cost of quantity as a result of a change in quality. The cost cross effect is positive when the marginal cost of quantity increases and negative when it decreases. Earlier papers have modelled the relationship between quantity and quality in the patient's utility but not in cost of production (Ma & McGuire, 1997), cost cross effect when quality is a positive function of quantity (Kesteloot & Voet, 1997), and no relationship between quality and quantity (Ma, 1998; Chalkley & Malcomson, 1995; Schleifer, 1985). This paper's contribution is to show the conditions under which quality can be induced when quantity and quality are complements or substitutes, and when quality is cost-increasing or cost-decreasing under three payment schemes: prospective payment, fixed fee for service, and cost-sharing.

Under prospective payment, the reimburser makes a fixed payment to the provider regardless of the quantity and quality of services produced. The reimburser, under the fixed fee for service, pays a fixed fee for each unit of quantity of service. Cost-sharing is a combination of prospective payment and full-cost reimbursement[32] and involves the provider receiving a fixed payment regardless of the quantity of services provided (prospective payment) and an additional fee that covers part of the marginal cost of providing the service (full-cost reimbursement). Below I summarize the literature relevant for the analysis.

In Ma and McGuire (1997), quality and quantity are related in the patient's utility but separable in the cost function. They show that when quantity and quality are substitutes,[33] high levels of quality are only possible if the payment at the margin is less than cost. Thus negative payment at the margin induces higher quality than positive payment as under cost sharing or zero payment as under prospective payment. However, a negative payment will induce the physician to underreport quantity. Truthful reporting of quantity then requires the prospective payment scheme but the resulting quality is too little. When quantity and quality effort are complements, the efficient effort is supplied only if the provider receives a payment that is strictly greater than the marginal cost of quantity. This is a requirement that prospective payment and cost sharing do not meet and so are ineffective in inducing the efficient effort.

In examining the incentive effects of reimbursement schemes to hospitals, Kesteloot and Voet (1997) take into account the possibility of co-operative agreements as well as quality and quantity competition among hospitals when quality is a positive function of quantity and cost cross effect is positive or negative. Their results show that whether hospitals compete or co-operate, an increase in the fee for service has a positive effect on quality. Such positive effect is higher the larger the cost-reducing and the quantity increasing effect of quality. An increase in the fee raises revenue especially if quantity is greater but increasing quantity also raises cost. Thus profit increases if higher quality boosts quantity and reduces cost.

Schleifer (1985) has both quantity and quality effort but does not relate these in terms of cost or demand. He finds that firms produce the efficient effort as long as the payment received is independent of cost, hence making the fixed fee for service the payment scheme that induces efficient effort. He shows that if firms are identical in cost and allowed to compete, then each firm will have the incentive to minimize cost if all firms are paid the same fee.

Ma (1998) models only quality and so has no relationship between quantity and quality. He distinguishes between effort that enhances quality and that reduces cost. Both types of efforts impose disutility on the altruistic provider and are both unobservable to the reimburser. He

---

[32] Under full cost reimbursement the reimburser pays an amount equal to the cost of provision of care. This payment scheme is becoming rare and so will not be covered here.

[33] Substitute and complements are used to refer to gross substitutes and complements in the patient's utility.

computes the level of fee that induces the optimal quality and effort when the provider's profit is positive and when it is zero. The results show that when the degree of altruism is small then the provider will only implement the optimal quality effort if it makes strictly positive profit. A higher degree of altruism will induce the optimal effort when profit is zero. The reimburser then has to set the level of prospective payment and cost margin (in the case of full-cost reimbursement) that induces the optimal quality effort.

Other papers have also examined how the reimbursement schemes affect quality. Fixed fee for service, which is often referred to in the literature as prospective payment because of the fixed nature of the fee, has been shown to provide incentive for efficient management and induce quality improvement (Gertler, 1989; Lee et al., 1983; Nyman, 1985) and reward providers for quality improvement if the quality improvement is cost reducing (Schleifer, 1985; Kesteloot & Voet, 1998; Wyszewianski et al. 1987; Norton, 1992). The level of the fee has been shown to be important in inducing quality improvement (Ma, 1998; Kooteloot & Voet, 1997; Chalkley & Malcomson, 1995; Norton, 1992; Hanchak et al., 1996). Kosecoff et al., (1990), found that patients are discharged sooner (decrease in quantity) under the prospective payment than under the full-cost reimbursement and this increases mortality rate as well as readmission rate (decrease in quality). Cost sharing is effective in inducing the desired quality if the optimal fraction of cost is covered (Keeler, 1990; Pope, 1990; Laffont & Tirole, 1987; Ellis & McGuire, 1986; Rickman & McGuire, 1999). Keeler, Pope, and Laffont &Tirole model their cost as a function of quality and effort but not quantity.

Even though the literature provides insights on the effect of different payment schemes on quantity and quality of care, it has not examined the combined effect of the relationship between quantity and quality in cost and the patient's utility. The role of the relationship between quality and quantity in cost shows the extent to which the choice of quality affects profit. Since quantity of service changes with payment scheme, it is reasonable for the provider to take the effect of change in quality on the marginal cost of quantity (the cost cross effect) into account when choosing quality. Ignoring the cost cross effect then reduces a model's ability to explain fully why quality is higher under some payment schemes than others. For example if a provider purchases an MRI machine, then it may enhance accuracy in diagnoses, reflecting quality improvement. However, the operation of the machine may require expertise and the use of costly accessories, thus increasing the marginal cost of quantity of service (positive cost cross effect). The provider may still find it profitable to purchase the machine if it can retain any cost saving that results from reducing the quantity of services even though the cost per service is high at the margin; this makes prospective payment an effective way to induce quality increase.

The relationship between quality and quantity in the patient's utility incorporates the effect of change in the quality or quantity on the patient's welfare and hence the demand for the provider's services. Using the example of the MRI machine above, the provider's incentive to purchase the machine may increase when quantity and quality are substitutes such that reducing quantity and increasing quality may not make the patient worse off and so the provider does not lose patients. This adjustment of quality and quantity to keep the patient at a level of utility is not captured in Kesteloot and Voet, where an increase in quality is necessarily accompanied by an increase in the aggregate quantity of services demanded. Ma and McGuire (1997) are also able to capture such effect since they model relationship between quality and quantity in the patient's utility. They do not, however, consider cost cross effect.

The literature has also shown the importance of altruism in providing quality and quantity (e.g., Ellis and McGuire, 1986; Kesteloot & Voet, 1997; Ma, 1998). Even in Kesteloot & Voet, where the patient could observe quality and the hospitals compete in quality, the provider supplied a higher level of quality when altruistic. Where cost cross effect and relationship in utility have an opposing effect on quality, the degree of altruism of the provider will determine which relationship should dominate. In the MRI example above, if quantity and quality are complements and so purchasing the machine would require increasing quantity of services and the marginal cost of quantity, the provider will purchase the machine as long as she puts greater weight on patient's welfare than on profit. It is thus important that these three factors be combined when examining the effect of payment schemes on quality.

This paper models the behaviour of an altruistic provider when quality is not contractible and so the provider is reimbursed for quantity. Taking cost effect into account, the paper compares the level of quality under the three payment schemes when quantity and quality are complements, gross substitutes and net substitutes. The paper presents a sequential game of a physician and a reimburser who is also the government. The reimburser chooses the payment scheme, and the physician chooses the quantity and quality to treat the patient. The reimburser maximizes the patient's utility subject to a participation constraint for the physician. The physician cares about the patient as well as profit, and so takes the patient's utility and cost into account when choosing quantity and quality.

The results show that the combined relationship between quality and quantity in the patient's utility as well as in cost plays a very important role in determining the level of quality and quantity supplied under any payment scheme. The relationship between quantity and quality in utility alone is not enough to determine the effect of payment scheme on quality. The cost cross effect of quality is also required to determine the effect of payment scheme on quality. The inclusion of such effect allows the paper to show the conditions under which financial incentives can induce or reduce quality. The results also show that unlike cost sharing, there is no condition under which fixed fee for service or prospective payment achieves first best results. For efficiency, however the reimburser needs to cover at least half of the cost of care.

The paper is divided into five sections. Section 2 describes the model, while Sections 3 derives and compares the equilibrium quality and quantity under the three payment schemes. Section 4 discusses the reimburser's strategies to induce an efficient supply under the three payment schemes while Section 5 concludes.

2.2 The Model

I model the relationship between the reimburser and the physician as a two-stage game. In the first stage, the reimburser chooses the reimbursement scheme that maximizes social surplus. In the second stage, the physician chooses the quality and quantity of care for the patient, given the reimbursement scheme. As in Ellis and McGuire (1986), the patient is passive; he simply accepts the quantity and quality of care the physician chooses for him. The patient is fully insured, gets sick only once in the period, and, as in Ma and McGuire (1997), derives benefit (or utility), $B(q, s)$, from care, where q represents the quantity of care and s represents quality of care.[34] I define quantity of care as the number of visits to the physician's office, or length of stay at a hospital. I define quality of care as the effectiveness of treatment in improving health. This includes both the physician's skill and all that is required to increase the patient's comfort. The patient has a well-behaved utility function $B_q > 0$, $B_s > 0$, and $B_{qq} < 0$, $B_{ss} < 0$ and quantity and quality are substitutes, $B_{qs} < 0$ or net substitutes[35] or $B_{qs} = 0$, or as complements, $B_{qs} > 0$.

Like Ellis and McGuire (1986), I assume that the physician is altruistic and knows how quantity and quality of care affect the patient's utility. The physician then takes the patient's utility into account when making her choices. She also cares about profit. As in Ma and McGuire (1997), the physician receives payment for quantity of care, but not for quality because quality is not contractible. The physician incurs cost, $c(q, s)$, with $c_i > 0$, $c_{ii} > 0$, $i = s, q$. As in Kesteloot and

[34] The fully insured patient means that the patient does not pay any fee to the physician. This implies that the physician receives payment only from the reimburser and nothing from the patient.
[35] This refers to the case in which trading off quantity for quality keeps the patient on the same level of utility irrespective of the level of quantity or quality.

Voet (1998), the cross partials of the cost function[36], c_{sq}, can be positive or negative. The physician's profit is $\pi = R(q) - c(q, s)$ where $R(q)$ represents the payment that the physician receives for providing treatment.

The physician cares about profit as well as the benefit that the patient receives from treatment and so her utility function is $U(B(q, s), \pi(q, s)) = \theta B(q, s) + \gamma\pi(q, s)$[37], where $0 < \theta < 1$, $\gamma = (1 - \theta)$ are the weights that the physician puts on the patient's benefit and profit respectively. As in Ellis and McGuire (1986), the marginal rate of substitution between patient utility and profit, $\alpha = \theta/\gamma$, indicates the degree of altruism. The physician is perfectly altruistic if she puts equal weight on patient benefit and profit, ($\alpha = 1$)[38]. The physician is less altruistic if she puts more weight on profit than on patient benefit, $\alpha < 1$. In the real world the physician is more likely to put a greater weight on profit than patient's utility (Ellis and McGuire, 1986), and so I do not examine the case in which the physician puts more weight on patient's benefits than profit, $\alpha > 1$. The physician chooses q and s to maximize her utility under each reimbursement scheme.

There are three reimbursement schemes: fixed fee for service, prospective payment, and cost sharing. Under fee for service, the physician receives a fixed fee, p, for each unit of quantity of care provided so that $R(q) = pq$. Under prospective payment, the physician receives a fixed payment, G, for the period regardless of quantity and quality of care provided, implying that $R(q) = G$. The cost-sharing scheme involves a combination of prospective payment and full cost reimbursement: the reimburser pays a fraction, τ, of the physician's cost in addition to a fixed payment, G; hence, $R(q) = G + \tau c(q, s)$. Given the above, the physician's profit can be written in a more general form as $\pi = G + pq - (1 - \tau)c(q, s)$ where $G > 0$ and $p = \tau = 0$ under prospective payment, $G = \tau = 0$ and $p > 0$ under fixed fee for service, and $G > 0$, $\tau > 0$ and $p = 0$ under cost sharing.

The reimburser's objective is to choose the payment scheme that would induce the physician to provide the socially efficient level of quantity and quality of care. Following Ma (1998), I model the regulator's preferences as maximizing the patient's utility $[B(q, s) - R(q)]$, subject to a participation constraint for the physician $[\theta B(q, s) + \gamma(R(q) - c(q, s)) = \bar{u}]$. The reimburser's objective function then is: $(1 + \alpha)B(q, s) - c(q, s) - \bar{u}/\gamma$.[39] I use backward induction

[36] The cross partials are referred to in the current paper as cost-cross effect.
[37] This form of utility function is the similar to Ellis and McGuire, (1986).
[38] Ellis and McGuire refer to the physician in this case as a perfect agent.
[39] From the constraint, $R(q) = u/\gamma - \alpha B(q, s) + c(q, s)$ which is substituted into the objective function to get: $(1 + \alpha)B(q, s) - c(q, s) - u/\gamma$.

to solve the game. The physician first chooses q and s taking the reimbursement scheme as given. The reimburser then chooses the reimbursement scheme that induces the efficient supply of q and s. The last stage of this game is analyzed in Section 3 while the first stage is investigated in Section 4.

2.3 The Physician

The physician's utility is $U(B(q, s), \pi(q, s)) = \theta B(q, s) + \gamma\pi(q, s)$, where $\pi = G + pq - (1-\tau)c(q,s)$ is her profit. The physician then about the patient as well as profit and so her utility increases with the patient's utility and/or profit. The first-order conditions of the physician's problem using the general form of profit are:

$$\frac{\partial U}{\partial q} = \alpha B_q(q,s) + p - (1-\tau)c_q(q,s) = 0 \qquad \textbf{(1)}$$

$$\frac{\partial U}{\partial s} = \alpha B_s(q,s) - (1-\tau)c_s(q,s) = 0 \qquad \textbf{(2)}$$

where $\alpha = \theta/\gamma > 0$. To ensure a maximum, I assume that the second-order conditions are satisfied.[40] Since I want to examine and compare the equilibrium s and q under each payment scheme, I need tools to do so. The simplest way is to use graphs in (s, q) space and graph the two first-order conditions above.

The slopes of $U_q = U_s = 0$ in (s, q) space are as follows:

$$\frac{\partial s}{\partial q}|_{U_q=0} = -\frac{\alpha B_{qq} - (1-\tau)c_{qq}}{\alpha B_{qs} - (1-\tau)c_{qs}} \quad \text{and} \quad \frac{\partial s}{\partial q}|_{U_s=0} = -\frac{\alpha B_{qs} - (1-\tau)c_{qs}}{\alpha B_{ss} - (1-\tau)c_{ss}}, \qquad \textbf{(3)}$$

These slopes can either be both negative or both positive. There is no condition under which one curve is positive and the other negative. To see this, observe first that the restrictions imposed on $B(s, q)$ and $c(s, q)$, $B_s > 0$, $B_q > 0$, $B_{ss} < 0$, $B_{qq} < 0$, $c_s > 0$, $c_q > 0$, $c_{ss} > 0$, $c_{qq} > 0$, make the numerator of $\frac{\partial s}{\partial q}|_{U_q=0}$ and the denominator of $\frac{\partial s}{\partial q}|_{U_s=0}$ both negative. Second, the denominator of $\frac{\partial s}{\partial q}|_{U_q=0}$ and the numerator of $\frac{\partial s}{\partial q}|_{U_s=0}$ are identical so they have the same signs.

[40] The second order conditions are: $H_1 = (\alpha B_{qq} - (1-\tau)c_{qq}) < 0$, $H_2 = H = (\alpha B_{ss} - (1-\tau)c_{ss})(\alpha B_{qq} - (1-\tau)c_{qq}) - (\alpha B_{sq} - (1-\tau)c_{sq})^2 > 0$ implying the Hessian is negative definite.

Thus, if under some conditions the denominator of $\frac{\partial s}{\partial q}|_{U_q=0}$ is negative (positive) then under the same conditions the numerator of $\frac{\partial s}{\partial q}|_{U_s=0}$ is also negative (positive) and both curves will be negatively (positively) sloped. From the second-order conditions, $\frac{\partial s}{\partial q}|_{U_q=0}$ is steeper than $\frac{\partial s}{\partial q}|_{U_s=0}$ and the equilibrium is at the point where the two curves cross.[41] The second derivative and the accompanying assumptions[42] show that the curves are convex. I will compare the equilibrium s and q under the different payment schemes in two categories of cases. The first category of cases (called hereafter Cases I) refers to the cases in which $U_q = U_s = 0$ are negatively sloped in (s, q) space while the second category of cases (called hereafter Cases II) refers to cases in which they are positively sloped. Cases I requires either $B_{qs} < 0$ and $c_{sq} > 0$; $B_{qs} < 0$ and $c_{sq} < 0$ but $|B_{qs}| > |c_{sq}|$, or $B_{qs} > 0$ and $c_{sq} > 0$ but $|B_{qs}| < |c_{sq}|$, while Cases II requires either $B_{qs} > 0$ and $c_{sq} < 0$; $B_{qs} > 0$ and $c_{sq} > 0$ but $|B_{qs}| > |c_{sq}|$, or $B_{qs} < 0$ and $c_{sq} < 0$ but $|B_{qs}| < |c_{sq}|$. *Figure 1a* illustrates Cases I and *Figure 1b* illustrates Cases II.

Since U_q shows different levels of q that the physician chooses to maximize her objective function given different levels of s, and since U_s shows that different levels of s she chooses given different levels of q, the optimal choice of the physician is found at the intersection of these two curves. Naturally, when U_q shifts to the right more q is provided, for each level of s, and when it

---

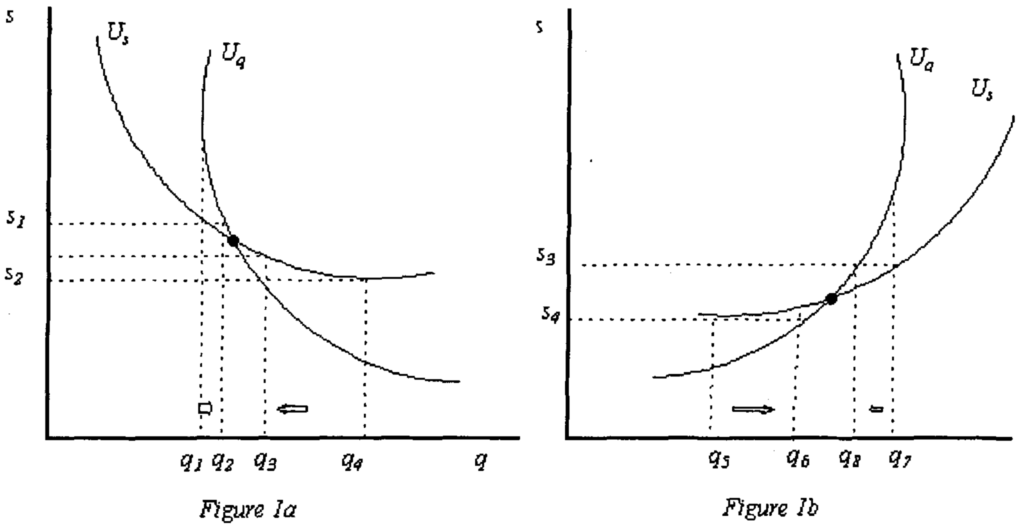
[41] Rearranging $H = (\alpha B_{ss} - (1 - \tau)\, c_{ss})(\alpha B_{qq} - (1 - \tau)\, c_{qq}) - (\alpha B_{sq} - (1 - \tau)\, c_{sq})^2 > 0$ produces $\frac{\alpha B_{qq} - (1-\tau)c_{qq}}{\alpha B_{qs} - (1-\tau)c_{qs}} > \frac{\alpha B_{qs} - (1-\tau)c_{qs}}{\alpha B_{ss} - (1-\tau)c_{ss}}$ if and only if $(\alpha B_{sq} - (1 - \tau)\, c_{sq}) < 0$ and if and $\frac{\alpha B_{qq} - (1-\tau)c_{qq}}{\alpha B_{qs} - (1-\tau)c_{qs}} < \frac{\alpha B_{qs} - (1-\tau)c_{qs}}{\alpha B_{ss} - (1-\tau)c_{ss}}$ only if $(\alpha B_{sq} - (1 - \tau)\, c_{sq}) > 0$.

[42] $\frac{\partial^2 s}{\partial q^2}|_{U_q=0} = \frac{(\alpha B_{sqs} - (1-\tau)c_{sqs})\left(\frac{\partial s}{\partial q}\right)^2 - (\alpha B_{qqq} - (1-\tau)c_{qqq})}{\alpha B_{qs} - (1-\tau)c_{sq}} > 0.$ The sign results from the assumptions that $(\alpha B_{sq} - (1 - \tau)\, c_{sq})$ and $\alpha B_{sqs} - (1 - \tau)c_{sqs}$ have the same signs and $(\alpha B_{sqs} - (1 - \tau)c_{sqs}) \left(\frac{\partial s}{\partial q}\right)^2 > \alpha B_{qqq} - (1 - \tau)c_{qqq}$. Note that $\alpha B_{sqs} - (1 - \tau)c_{sqs}$ has no interesting economic interpretation and so its sign and magnitude do not affect the analysis. The second derivative of $U_s = 0$ is the same as that of $U_q = 0$ except that $\alpha B_{sqs} - (1 - \tau)c_{sqs}$ becomes $\alpha B_{sss} - (1 - \tau)c_{sss}$, $\alpha B_{qqq} - (1 - \tau)c_{qqq}$ becomes $\alpha B_{sqq} - (1 - \tau)c_{sqq}$, and the denominator becomes $\alpha B_{ss} - (1 - \tau)c_{ss}$.

shifts to the left less q is provided. When U_s shifts up more s is provided for each level of q and when it shifts downward less s is provided.

Observe that under Cases I and II, the equilibrium is stable in the sense that a small deviation from it brings the physician's optimal choice of q and s back to the intersection. To see this, consider q_1 in *Figure 1a*. According to U_s, s_1 is then the optimal choice. However, when s_1 is chosen, the corresponding optimal q (according to U_q) is q_2. Hence any perturbation taking q or s outside the intersection brings them back to the intersection. It is straightforward to see that the same reasoning applies when the perturbation shifts s or q on the other side of the point of intersection in *Figure 1a* or when the curves are upward sloping as in *Figure 1b*.

Figure 2.1: Stability of Equilibrium



The economic intuition for these two cases is the following. Cases I represent the different combinations of cost effects and the patient's preferences and trade-off between quantity and quality of care such that it is utility-maximizing for the physician to increase s, respectively q, when she decreases q, respectively, s. This occurs when s and q are substitutes and the cost-cross effect is positive ($B_{qs} < 0$ and $c_{sq} > 0$), or negative but small ($B_{qs} < 0$ and $c_{sq} < 0$ but $|B_{qs}| > |c_{sq}|$) or else when and q are complements and the cost cross effect is positive and strong ($B_{qs} > 0$ and $c_{sq} > 0$ but $|B_{qs}| < |c_{sq}|$).

Consider the first possibility. When q and s are substitutes, then a fall in q should be accompanied by an increase in s otherwise the patient's utility falls and *ceteris paribus,* a fall in the patient's utility decreases the physician's utility. When the cost-cross effect is positive, a decrease in q decreases the marginal cost of s and so it is profit increasing to increase s. Thus when q and s are substitutes and the cost cross effect is positive, increasing s when q has decreased, increases both profit and the patient's utility and so increases the physician's utility.

Consider now the second of these possibilities. When s and q are complements increasing s and decreasing q decreases the patient's utility; however, when the cost cross effect is positive and strong enough, the resulting increase in profit outweighs the fall in the patient's utility, and so the physician's utility increases. Negative cost cross effect implies that a decrease in q increases the marginal cost of s, and so it is profit-decreasing to increase s. However, when this effect is small and q and s are substitutes, then increasing s when q is decreased results in an increase in the physician's utility because the increase in the patient's utility outweighs the fall in profit. These cases will explain why certain payment schemes cause s and q to change in the opposite directions.

Cases II collect cases in which $U_q = U_s = 0$ is positively sloped in (s, q) space. This corresponds to the combinations of cost effects, patient's preferences and trade off concerning quantity and quality of care such that it is utility-maximizing for the physician to increase s, respectively q, when she increases q, respectively s. This occurs either when s and q are substitutes and the cost cross effect is negative and strong ($B_{qs} < 0$ and $c_{sq} < 0$ but $|B_{qs}| < |c_{sq}|$), or else when s and q are complements and the cost cross effect is positive but small ($B_{qs} > 0$ and $c_{sq} > 0$ but $|B_{qs}| < |c_{sq}|$), or negative ($B_{qs} > 0$ and $c_{sq} < 0$).

When the cost cross effect is negative, increasing s when q has risen increases profit. When q and s are complements, increasing s when q has been increased leads to an increase in the patient's utility and, combined with negative cost cross effect, reinforces the increase in the physician's utility. A positive but small cost cross effect, given s and q are complements, implies that any fall in profit resulting from increasing q and s is outweighed by the increase in the patient's utility. Again this category of cases will explain why the physician increases both s and q under some payment schemes. As will be shown later these two cases are important in explaining the different levels of equilibrium q and s under different payment schemes.

Table 2.1: Examples of Substitutes, Complements and Cost Cross Effects

Relationship	Example
Substitutes	Minor Surgeries. By increasing monitoring (increase in s), q could be reduced.
Complements	Major Surgeries. A by-pass requires intensive care (high s) and long length of stay
Positive Cost Cross Effect	An introduction of a new technology to improve the quality of treatment (increase in s) may require more time of the physician to study a given case to be able to match treatment with illness.
Negative Cost Cross Effect	The use of machines that allow patients to operate killer or take their own blood pressure and so requires less care provider's time.

2.3.1 Fixed Fee for Service

The first-order conditions under this reimbursement scheme are **(1)** and **(2)** above with $\tau = 0$. Denote the solutions under fixed fee for services as $q^F(\alpha, p)$ and $s^F(\alpha, p)$. To examine how the level of the fixed fee, p, affects the physician's choice of quantity, and to compare with quality and quantity under the other payment schemes, I do comparative statics with respect to p,

$$\frac{\partial q^F}{\partial p} = -\frac{\alpha B_{ss} - c_{ss}}{H} > 0 \tag{4}$$

$$\frac{\partial s^F}{\partial p} = \frac{\alpha B_{sq} - c_{sq}}{H} < 0 \text{ under Cases I and} > 0 \text{ under Cases II} \tag{5}$$

where $H > 0$ represents the Hessian,[43] and where $B_{ss} < 0$, $c_{ss} > 0$. Equation **(4)** shows that the quantity of services always increases when the level of the fee for services increases. This is not surprising because p is the fee that the physician receives for supplying q, and so **(4)** indicates

[43] Recall that $H = (\alpha B_{ss} - c_{ss})(\alpha B_{qq} - c_{qq}) - (\alpha B_{sq} - c_{sq})^2 > 0$

an upward sloping supply curve. From **(1)** $U_{qp} > 0$, so an increase in p requires a higher level of q given any s for the first-order condition $U_q = 0$ (this is the case because $B_{qq} < 0$ and $c_{qq} > 0$). In other words, to a higher p is associated a curve $U_q = 0$ further to the right.

However, **(5)** shows that the effect of a change in the fee on quality of care depends on the cases, i.e., whether quantity and quality are substitutes, $B_{sq} < 0$, net substitutes, $B_{sq} = 0$, or complements, $B_{sq} > 0$, and whether the cost cross effect is positive, $c_{sq} > 0$, or negative, $c_{sq} < 0$. The numerator of **(5)** is the same as both the denominator of $\frac{\partial s}{\partial q}|_{U_q=0}$ and the numerator of $\frac{\partial s}{\partial q}|_{U_s=0}$. Thus a negative sign in **(5)** corresponds to Cases I, while a positive sign in **(5)** corresponds to Cases II. Equation **(5)** shows that under Cases I an increase in the level of the fee leads to a reduction in the equilibrium s. Since $U_{sp} = 0$, U_s does not shift in *Figure 2a* and the equilibrium moves from F to F'.

The intuition for such results is the following. When s and q are substitutes, quality decreases with p when the cost cross effect is positive. Intuitively, the positive cost cross effect implies that increasing quality would increase the marginal cost of quantity of care. Since the increase in the fee also increases quantity, increasing quality would have an adverse effect on profit through higher cost. Thus quality should decrease with the fee when the cost cross effect is positive. Moreover, lower quality and higher quantity may not make the patient worse off when the two are substitutes.[44] A decrease in quality in this case may increase both the patient's utility and profit and so be consistent with the maximization of the physician's objective function.

[44] This is based on the assumption that the relationship between quantity and quality are such that the fall of quality and the rise in quantity here does not make the patient worse off. Even if the decrease in quality makes the patient worse off, the increase in profit is high enough to increase the physician's utility.

Figure 2.2: Effect of the Fee on Quality and Quantity

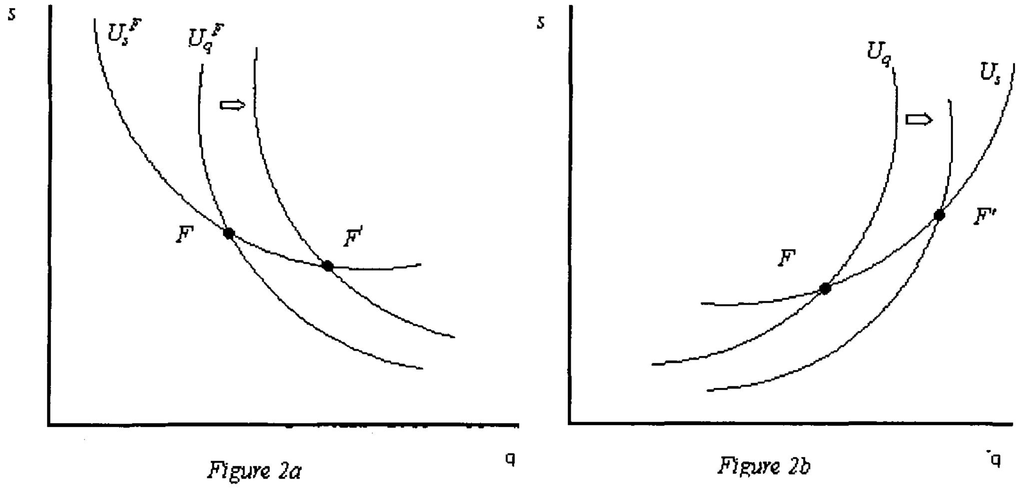


Figure 2a *Figure 2b*

Equation (5) also shows that quality increases with p only under Cases II. As illustrated in *Figure 2b*, this corresponds to a shift of the equilibrium from F to F', which is characterized by higher quantity and quality. To compare such results with those in the literature I will examine the intuition for each case under Cases II. First, when quantity and quality are complements and the cost cross effect is positive, an increase in the fee for service will increase quality only if the cost cross effect is not strong. Intuitively, when s and q are complements, then a higher quality when quantity has increased, resulting from an increase in the fee, makes the patient better off. However, the positive cost cross effect implies that increasing quality could decrease profit. The provider then will increase quality if the improvement in the patient's utility outweighs the decrease in profit.

Second, when s and q are complements and the cost cross effect is negative, quality always increases with the fee. Intuitively, a negative cost cross effect implies that an increase in quality decreases the marginal cost of quantity and so increasing quality is profit-improving. Quality increase under such conditions is both profit and patient utility improving and so is consistent with the physician's maximization. This result is similar to Kesteloot and Voet (1998) in that the negative cost cross effect induces an increase in quality. However, the current model shows that the negative cost cross effect is only part of the story. The complementarity of quality and quantity also explains the increase in quality with an increase in the fee when the cost cross effect is negative.

Note that when the relationship between quality and quantity in the patient's utility are not considered, i.e., $B_{sq}(q, s) = 0$, then the numerator in **(5)** is equal to $-c_{sq}$ and so, as in Kesteloot & Voet, quality will always fall with the fee when the cost cross effect is positive. Thus by adding the relationship between quality and quantity in the patient's utility, I am able to show that the positive cost cross effect causes quality to decrease from an increase in the fee only when quantity and quality are substitutes or when they are complements but the cost cross effect dominates. When quality and quantity are complements, quality can increase with the fee even when the cost cross effect is positive as long as the resulting increase of the patient's utility is significant relative to the fall in profit.

Third, when quantity and quality are substitutes and the cost cross effect is negative, quality will only increase with the fee when the negative cost cross effect is stronger than the reduction in the patient's utility. Otherwise, quality will decrease with the fee even when the cost cross effect is negative. Intuitively, quantity increases with the fee and, since quality and quantity are substitutes, decreasing quality may not make the patient worse off. However the negative cost cross effect improves profit when quality increases. The physician then increases quality (as in *Figure 2b*) only if the negative cost cross effect is significant enough to increase profit, otherwise he decreases quality (as in *Figure 2a*). This is contrary to Kesteloot and Voet, where negative cost cross effect always induces quality increase under the fixed fee for service. This is not surprising because Kesteloot and Voet do not consider the case in which s and q are substitutes and so there is no force in their model working against the negative cost cross effect when the fee increases. The main results of the current model under fixed fee for service that are not found in the existing literature are summarized below:

PROPOSITION 1: *An increase in the fee decreases quality and increases quantity under Cases I where quality and quantity are substitutes and the cost cross effect is either positive or weak when negative, or when quality and quantity are complements and the positive cost cross effect is strong. Under Cases II, where quality and quantity are complements and the cost cross effect is negative or when it is positive but weak, an increase in the fee increases quality and quantity.*

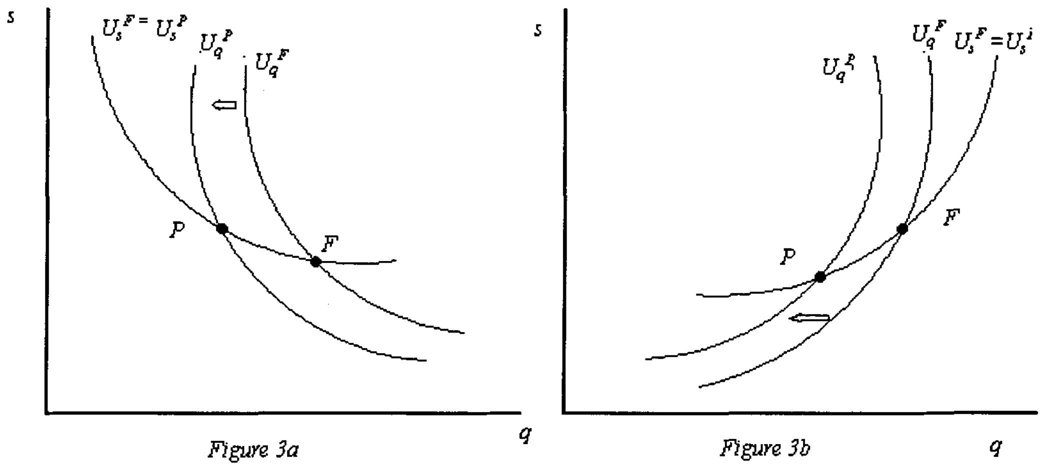
2.3.2 Prospective Payment

The first-order conditions under prospective payment is again given by **(1)** and **(2)**, but with $p = 0$ and $\tau = 0$. The only difference between the first-order conditions in the present case and those of the fixed fee for service is thus $p = 0$. The equilibrium quantity and quality under prospective payment are $q^P(\alpha)$ and $s^P(\alpha)$. Hence, all things being equal, $q^F(\alpha, p)$ and $s^F(\alpha, p)$ should approach $q^P(\alpha)$ and $s^P(\alpha)$ as p approaches zero. Note that $q^P(\alpha)$ and $s^P(\alpha)$ are both zero when $\alpha = 0$ and so $q^P(\alpha)$ and $s^P(\alpha)$ are positive only when the physician is altruistic, i.e., $\alpha > 0$. The analysis focuses on the case in which the physician is altruistic.

Quantity and quality of care under fee for service would exceed those under prospective payment if p has a positive effect on q and s. Let U_q^P and U_s^P denote the $U_q = 0$ and $U_s = 0$ under the prospective payment. Equation **(1)** shows that while $U_{qp} > 0$ under the fee for service, $U_{qp} = 0$ under prospective payment. It follows that given s, $q^P < q^F$ and thus a switch from fee for service to prospective payment shifts the U_q to the left. However, from **(2)**, $U_{sp} = 0$ under both payment schemes, and so $U_s^F = U_s^P$. Assuming the equilibrium under the fixed fee for service is F in *Figure 3*, then the equilibrium under prospective payment is P. *Figure 3a* and *3b* show that $q^P(\alpha) < q^F(\alpha, p)$ but the difference between $s^P(\alpha)$ and $s^F(\alpha, p)$ depends on the cases.

Figure 3a shows that under Cases I a switch from fixed fee for service to prospective payment decreases quantity and increases quality, $q^P(\alpha) < q^F(\alpha, p)$ and $s^P(\alpha) > s^F(\alpha, p)$. The lower q and higher s resulting from switching from fee for service to prospective payment is consistent with the negative relationship between q and s under this category of cases.

Figure 2.3: Comparing the Levels of Quality and Quantity under the Fixed Fee for Service and Prospective Payment



The increase in quality in this case, however, is not necessarily utility improving for the patient because prospective payment decreases quantity, and so increasing quality, when quantity and quality are complements, does not make the patient better off. Thus, when profit dominates the physician's decision, quality improvement could make the patient worse off.

Figure 3b, shows that $q^P(\alpha) < q^F(\alpha, p)$ and $s^P(\alpha) < s^F(\alpha, p)$ under Cases II. The decrease in q and s from switching from fee for service to prospective payment is consistent with the positive relationship between s and q under this category of cases. The intuition for the fall of quality under Cases II comes from the cost cross effect. When the cost-cross effect is negative, the decreased q does not provide incentive to invest costly effort to improve quality in order to reduce the cost of already reduced q. A fall in quality in this case, then, is consistent with profit maximization.

However, whether quantity and quality are substitutes or complements, a switch from fixed fee for service to prospective payment makes the patient worse off under Cases II. When quality and quantity are complements, the patient is worse off because both quality and quantity decrease. When quantity and quality are substitutes, the patient is made worse off because a lowered quantity is accompanied by a lower quality. The low quality under Cases II then results when the physician puts more weight on profit than on the patient's health.

PROPOSTION 2: *A switch from fixed fee for service to prospective payment decreases quantity. Quality, however, increases under Cases I, but falls under Cases II. The patient may be worse off under Cases I but is definitely worse off under Cases II. The physician provides positive levels of quantity and quality only if she is altruistic.*

I now examine briefly the case of net substitutes. Equation (5) shows that when quantity and quality are net substitutes, $B_{sq} = 0$, quality increases with prospective payment (as in Cases I) when the cost cross effect is positive but decreases with prospective payment (as in Cases II) when the cost cross effect is negative. Intuitively, positive cost cross effect implies that an increase in quality increases the marginal cost of quantity, so the fall of quantity induced by prospective payment reflects efficiency gain to the provider in the form of cost reduction. Because marginal cost of quantity increases with quality, increasing quality effort per visit allows the physician to reduce the number of visits in order to reduce cost. A negative cost cross effect, however, implies that an increase in quality leads to a fall in the marginal cost of quantity. This kind of quality can take the form of better management of resources that leads to cost reduction. Equation (5) shows that this kind of quality falls with prospective payment when quantity and quality are net substitutes. Intuitively, it is cheaper for the physician to reduce quantity under prospective payment than to invest in quality that will reduce the marginal cost of quantity. Proposition 2 summarizes the findings:

PROPOSITION 3: *When quantity and quality are net substitutes, quality falls with prospective payment when the cost cross effect is negative but increases with prospective payment when the cost cross effect is positive.*

2.3.3 Cost Sharing

The first-order conditions under the cost-sharing scheme is the same as in **(1)** and **(2)** with $p = 0$, and $0 < \tau < 1$. I denote the equilibrium quantity and quality as $q^C(\alpha, \tau)$ and $s^C(\alpha, \tau)$ respectively. Recall that cost sharing is a combination of full cost reimbursement and prospective payment, so $\tau = 0$ and $G > 0$ represents prospective payment and $\tau = 1$ and $G = 0$ represents full-cost reimbursement. To compare $q^C(\alpha, \tau)$ and $s^C(\alpha, \tau)$ with those under the prospective payment, I do comparative statics with respect to τ:

$$\frac{\partial q^C}{\partial \tau} = -\frac{c_q(\alpha B_{ss} - (1-\tau)c_{ss})}{H} > 0 \qquad \textbf{(6)}$$

$$\frac{\partial s^C}{\partial \tau} = \frac{c_q(\alpha B_{sq} - (1-\tau)c_{sq})}{H} < 0 \text{ under Cases I and} > 0 \text{ under Cases II} \qquad \textbf{(7)}$$

With $\tau = 0$ under prospective payment, $q^C(\alpha, \tau)$ and $s^C(\alpha, \tau)$ should approach $q^P(\alpha)$ and $s^P(\alpha)$ respectively when τ approaches zero. Let U_q^C and U_s^C represent the $U_q = 0$ and $U_s = 0$ under cost sharing. With $U_{q\tau}^C > 0$ and $U_{q\tau}^P = 0$, $q^C > q^P$ given s where q^C (q^P) satisfies $U_q^C = 0$ ($U_q^P = 0$) i.e., a switch from prospective payment to cost sharing shifts U_q to the right.[45] When τ rises, then both U_q and U_s become positive. Since $B_{qq} < 0$ and $c_{qq} > 0$, then $U_s^C = 0$ requires a higher q for any given s. Similarly since $B_{ss} < 0$ and $c_{ss} > 0$, then $U_s^C = 0$ requires a higher s for any given q. This is shown in *Figure 4a* and *4b* with C as the equilibrium under cost sharing and P as that under prospective payment. Equation **(6)** shows that quantity increases with τ regardless of the cases i.e., q always increases as more cost is covered. Intuitively, when τ increases the physician bears less of the cost of providing quantity and so has the incentive to increase quantity. It follows that $q^C(\alpha, \tau) > q^P(\alpha)$ regardless of the cases.

[45] From **(1)**, $U_{q\tau} = c_q > 0$.

Figure 2.4: Comparing the Levels of Quality and Quantity under Prospective Payment and Cost Sharing

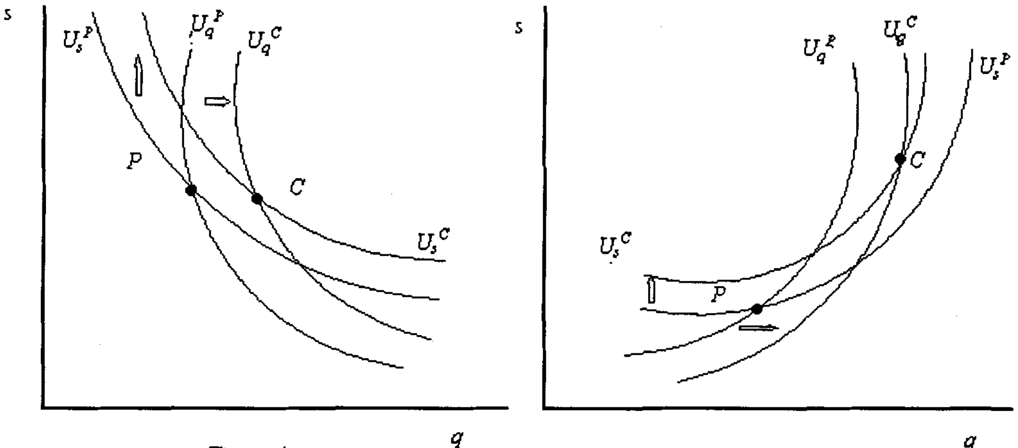


Figure 4a *Figure 4b*

Equation (7) shows that the effect of τ on quality depends on the cases. Under Cases I, an increase in τ leads to a decrease in s. It follows then from (7) that $s^C(\alpha,\tau) < s^P(\alpha)$. As shown in *Figure 4a*, the curves shift such that $q^C(\alpha,\tau) > q^P(\alpha)$ and $s^C(\alpha,\tau) < s^P(\alpha)$. The intuition is the following. When quality and quantity are substitutes and the cost cross effect is positive, quality falls with the cost coverage, τ. It is the case because, with quantity increasing with τ, $q^C(\alpha,\tau) > q^P(\alpha)$, reducing quality when quantity and quality are substitutes may not make the patient worse off. Besides, the positive cost cross effect would have adverse effect on profit if quality increases. Thus decreasing quality is consistent with maximization of the physician's utility.

Equation (7) also shows that under Cases II s increases with τ and so $q^C(\alpha,\tau) > q^P(\alpha)$ and $s^C(\alpha,\tau) > s^P(\alpha)$. The equilibrium is shown in *Figure 4b* as *C*. Again the intuition for the results depends on the cases. As an example the intuition behind the increase in quality with cost sharing when quality and quantity are complements and cost cross effect is negative or positive is as follows: With $q^C(\alpha,\tau) > q^P(\alpha)$, setting $s^C(\alpha,\tau) > s^P(\alpha)$ improves patient's health when quantity and quality are complements. The negative cost cross effect induces an increase in quality to reduce the cost of quantity and hence increase profit. Even though cost reduction incentive is still important under cost sharing, (7) shows that as the cost coverage, τ, increases, the cost cross effect, c_{sq}, becomes less important force in determining the level of quality. A full cost coverage, $\tau = 1$, then would eliminate the cost cross effect.

PROPOSITION 4: *Quantity under cost sharing exceeds that under prospective payment regardless of the category of cases. Under Cases I (Cases II) quality under cost sharing is less (greater) than that under prospective payment.*

The result is contrary to Ma and McGuire (1997) where cost sharing and prospective payment cannot induce an increase in quality when quantity and quality are complements. In their model $c_{sq} = 0$ and so only the relationship between quality and quantity in the patient's utility, B_{sq}, is considered. Equation (7) shows that when τ falls, quality will always fall if $c_{sq} = 0$ and $B_{sq} > 0$. However, when $c_{sq} > 0$ and this effect is strong, then quality will increase even if $B_{sq} > 0$. The inclusion of cost cross effect makes the current model able to show that cost sharing as well as prospective payment can induce quality improvement when quality and quantity are complements.

The results are, however, consistent with Kesteloot and Voet, who show that the incentive to increase quality under negative cost cross effect does not exist under the full cost reimbursement. Under the full-cost reimbursement the reimburser covers the total cost of production, and so the provider does not have the incentive to engage in quality improvement that reduces cost. The current model shows that cost sharing reduces the influence of the cost cross effect on the level of quality. In fact, the cost cross effect is eliminated under full cost reimbursement, $\tau = 1$. However, since the provider still bears some cost under cost sharing, the cost cross effect is not eliminated as under full cost reimbursement. I now compare the equilibrium choices under cost sharing with those under fixed fee for service.

Even though $\tau = 0$ under the fixed fee for service, $q^F(\alpha, p)$ and $s^F(\alpha, p)$ vary with p, and so the parameter τ alone cannot be used to compare $q^C(\alpha, \tau)$ and $s^C(\alpha, \tau)$ with those under fixed fee for service. However, from (2), $U_{s\tau}^C = c_s$, but $U_{s\tau}^F = 0$ so that a positive τ makes $U_s^C > 0$. Hence, for every given q and given $B_{ss} < 0$ and $c_{ss} > 0$, the physician supplies more s under cost sharing than under fixed fee for service. Equation (1) shows that when $\tau = \tau^C = \dfrac{p}{c_q(q,s)}$, the first-order condition under cost sharing is identical to that of fee for service[46] so $U_q^C = U_q^F$ for any given (s, q). Thus, for any given s the physician chooses the same level of q under both cost sharing and fixed fee for service. The resulting equilibrium, shown in *Figure 5* as C, has $q^C(\alpha, \tau)$

---

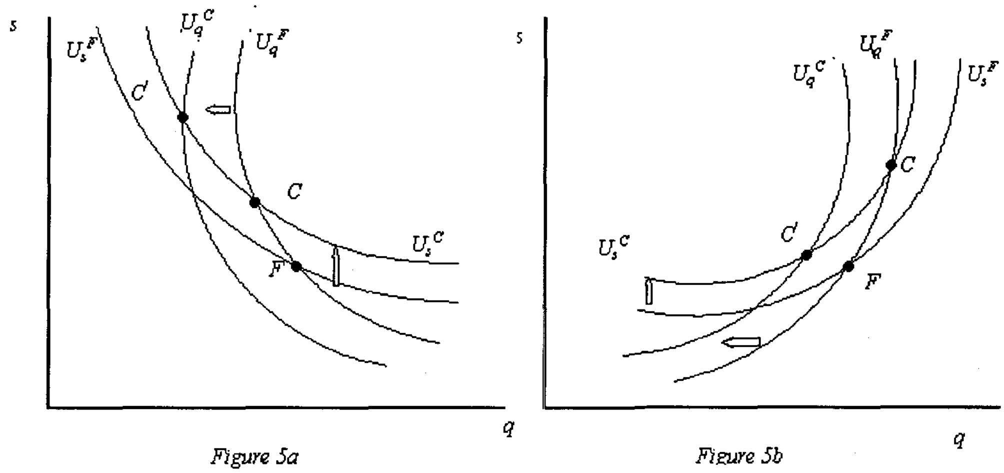
[46] It is easy to find that (1) for $p = 0$ and $\tau \neq 0$ becomes identical to the first-order condition in q with fixed fee once $\tau = \tau^C$. When $\tau = \tau^C$ then for a given s, $q^F = q^C$.

$< q^F(\alpha, p)$ and $s^C(\alpha,\tau) > s^F(\alpha, p)$ under Cases I and $q^C(\alpha,\tau) > q^F(\alpha, p)$ and $s^C(\alpha,\tau) > s^F(\alpha, p)$ under Cases II.

It is now easy to compare the equilibrium for $\tau \neq t^F$. For instance when $\tau < t^F$ then U_q^C is to the left of U_q^F for any given (s, q). To have $U_q^C = 0$, with $B_{qq} < 0$ and $c_{qq} > 0$, then for any given s the physician provides less q under cost sharing than fixed fee for service. In *Figure 5a*, when U_q^C is to the left of U_q^F then $q^C(\alpha,\tau) < q^F(\alpha, p)$ and $s^C(\alpha,\tau) > s^F(\alpha, p)$ under Cases I, shown as C'. Under Cases II $q^C(\alpha,\tau) \gtreqless q^F(\alpha, p)$ and $s^C(\alpha,\tau) > s^F(\alpha, p)$. *Figure 5b* is drawn for a case in which $q^C(\alpha,\tau) < q^F(\alpha, p)$. The reader can verify from *Figure 5b* that when τ is very small such that U_q^C is to the far left of U_q^F then $q^C(\alpha,\tau) < q^F(\alpha, p)$ and $s^C(\alpha,\tau) < s^F(\alpha, p)$ under Cases II.

When $\tau > t^F$, then $U_q^C > U_q^F$ i.e., in order for U_s^C to equal zero, given $B_{qq} < 0$ and $c_{qq} > 0$, then for any given s, more q is supplied under cost sharing than under fixed fee for service. The resulting equilibrium, not shown on the graph, has $q^C(\alpha,\tau) \gtreqless q^F(\alpha, p)$ and $s^C(\alpha,\tau) \gtreqless s^F(\alpha, p)$ under Cases I and $q^C(\alpha,\tau) > q^F(\alpha, p)$ and $s^C(\alpha,\tau) > s^F(\alpha, p)$ under Cases II. Even though the equilibrium is not shown on the graph the reader can verify in *Figure 5a* that $s^C(\alpha,\tau) < s^F(\alpha, p)$ is only possible under Cases I if U_q^C is to the further right of U_q^F and so $q^C(\alpha,\tau)$ is much greater than $q^F(\alpha, p)$.

Figure 2.5: Comparing the Levels of Quality and Quantity under Cost Sharing and Fee for Service



In summary, the relationship between the levels of quantity and quality under the two payment schemes depends on the size of τ and the cases. In general $s^C(\alpha,\tau) > s^F(\alpha, p)$ and $q^C(\alpha,\tau) < q^F(\alpha, p)$ under Cases I and $s^C(\alpha,\tau) > s^F(\alpha, p)$ under Cases II regardless of the relationship between $q^C(\alpha,\tau)$ and $q^F(\alpha, p)$ and the size of τ. However, when τ is much higher than τ^C, then $q^C(\alpha,\tau) > q^F(\alpha, p)$ and $s^C(\alpha,\tau) < s^F(\alpha, p)$ is possible under Cases I. Under Cases II, $q^C(\alpha,\tau) < q^F(\alpha, p)$ and $s^C(\alpha,\tau) < s^F(\alpha, p)$ is also possible when τ is very small relative to τ^C.

The intuition of the results comes from the effect of τ and p on quantity. From **(6)**, a high τ induces high levels of quantity. Thus, the physician supplies higher levels of quantity under cost sharing than under the fixed fee for service if τ is set above the critical level where quantity under the two payment schemes are equal. When τ is below the critical level such that the physician supplies less (or slightly greater) quantity under cost sharing than under fixed fee for service, she always supplies more quality under cost sharing than under fixed fee for service. The reason comes from the way quality is reimbursed. Under fixed fee for service the physician bears the full cost of quality but only bears part of the cost under cost sharing. Thus she has more incentive to supply more quality under cost sharing than under fixed fee, at least when she supplies less or almost the same quantity under both payment schemes.

When τ is high such that quantity under cost sharing is much greater than that under fixed fee, then under Cases I it is profit-maximizing for the physician to supply less quality under cost sharing than under fixed fee for service despite the partial coverage of cost under cost sharing. The intuition behind this is that when the cost-cross effect is positive and strong, a possibility under Cases I, increasing quality will increase the marginal cost of quantity. Given a large supply of quantity, increasing quality may reduce profit, and so quality has to decrease.

Under Cases II, quality under cost sharing exceeds that under fixed fee regardless of the relationship between quantities under the two payment schemes. The intuition is that when cost cross effect is negative and strong, for example, increasing quality will decrease the marginal cost of quantity. It is thus profit improving for the physician to increase quality whether quantity is high or low. However, when τ is close to zero then the incentive to increase quality falls. The following proposition results:

PROPOSITION 5: *As long as τ is not too different from the critical level, a switch from fixed fee for service to cost sharing provides the incentive to increase quality regardless of Cases I or II and the relationship between quantity under the two payment schemes. When τ is much greater (smaller) than the critical level then under Cases I (Cases II), quality under fixed fee for* *service exceeds that under cost sharing and quantity under cost sharing is much greater (smaller) than that under fixed fee for service.*

Proposition 5 is a new result with respect to the current literature. The literature so far has not compared quality (and/or quantity) under the fixed fee for service and cost sharing. For example, Ellis and McGuire (1997) as well as Rickman and McGuire (1999) compare quantities under cost sharing, prospective payment and full cost reimbursement. Kesteloot and Voet, (1998) Chalkey & Malcomson (1995) and Ma (1998) compare quality under fixed fee for service with that under full cost reimbursement. Ma and McGuire compared quality under prospective payment, with that under cost sharing. The current paper shows that quality under cost sharing can be greater or less than that under fixed fee for service.

To sum up, the results in *Section 3* show that in general, prospective payment induces the lowest provision of quantity while the relationship between quantity under fixed fee for service and cost sharing depends on the levels of p and τ. In the case of quality, prospective payment provides the highest level of quality under Cases I, and the lowest under Cases II. Quality under cost sharing always exceeds that under fixed fee for service regardless of the case and the difference in quantity as long as τ is not too different from the critical level. Fixed fee for service then provides the lowest quality under Cases I.

These results address the popular notion that allocating more resources to health care improves the quality of care provided. Whether or not allocation of more resources leads to increase in the quality of care depends on the cases. Increasing the fixed fee or cost sharing leads to a reduction of the quality of care under Cases I. For example when the fixed fee or cost sharing increases quantity increases but the negative relationship between quality and quantity leads to a reduction of quality. This negative effect of higher fee and cost sharing on quality comes from the non-contractibility of quality. Since contracts are based on quantity, the effect of the payment on quality depends on the relationship between quality and quantity. It is therefore possible to induce high levels of quality by increasing the payment to quantity as long as there is a positive relationship between quality and quantity. I now compare the equilibrium outcome with the efficient level of quality and quantity by examining the reimburser's strategies.

2.4 The Reimburser

As already explained, the reimburser's objective is to choose the reimbursement scheme that maximizes the objective function: $(1 + \alpha)B(q, s) - c(q, s) - \overline{u}/\gamma$. The reimburser's goal is thus to find a ranking of the various available schemes according to their ability to achieve efficiency. Under each payment scheme I examine the conditions under which the payment scheme can be used to induce the efficient provision of quantity and quality. The efficient quality and quantity are provided when the marginal benefit for the services equal the marginal cost of provision. The first order conditions that maximizes the social surplus are

$$(1+\alpha)B_q(q,s)-c_q(q,s)=0 \tag{8}$$

$$(1+\alpha)B_s(q,s)-c_s(q,s)=0 \tag{9}$$

The solutions to **(8)** and **(9)** are first-best and are denoted as q^* and s^*. Thus, q^* and s^* will provide a benchmark to compare q and s under the three payment schemes with regard to efficiency. The slopes of the schedules U_q^* and U_s^* are such that **(8)** and **(9)** are satisfied. They are the same as those in **(3)** but with $\alpha = (1+\alpha)$ and $\tau = 0$:

$$\frac{\partial s}{\partial q}\Big|_{U_q=0} = -\frac{(1+\alpha)B_{qq}-c_{qq}}{(1+\alpha)B_{qs}-c_{qs}} \quad \text{and} \quad \frac{\partial s}{\partial q}\Big|_{U_s=0} = -\frac{(1+\alpha)B_{qs}-c_{qs}}{(1+\alpha)B_{ss}-c_{ss}}, \tag{10}$$

Thus the slopes are negative under Cases I and positive under Cases II and are also convex.[47] I now compare this first-best solution to the market solution induced by each scheme. Under each payment scheme, I examine how the instruments under the payment scheme can be used to induce efficient supply of quality.

2.4.1 Fee For Service

Under the fixed fee for service, **(1)** shows that the level of p is important in determining the relationship between quality and quantity. The first-order condition determining q under fee for service, **(1)**, is identical to **(8)** as long as $p = p^* = B_q(q^*, s^*)$. This implies that if the reimburser can also choose p, then it chooses $p^* = B_q(q^*, s^*)$ in order to reach the first-best quantity. When $p > p^*$, **(4)** shows that quantity increases, so $q^F > q^*$. Similarly when $p < p^*$, then $q^F < q^*$.

In the case of quality, **(2)** shows that, given $q^F = q^*$, the physician undersupplies quality, $s^F < s^*$, even if she is perfectly altruistic, i.e., $\alpha = 1$.[48] Thus, it is not possible to use the fee to

[47] The second derivative is the same as those of **(3)** but with $\alpha = (1+\alpha)$ and $\tau = 0$.

[48] $(1+\alpha)B_s(q,s)-c_s(q,s)=0$ implies $B_s < c_s$ and $\alpha B_s(q,s)-c_s(q,s)=0$ implies $B_s \geq c_s$. The concavity assumptions then imply that $s^F < s^*$. [When $\alpha = 0$ **(2)** shows that the physician's marginal profit is $-c_s$, which is negative so that the minimum level of quality will be provided (i.e. $s^F = 0$)].

induce an efficient supply of both quantity and quality. An efficient supply of quantity is necessarily accompanied by a low supply of quality. This implies that the fee for service is an appropriate scheme only when the government is concerned about inducing the efficient supply of quantity but not quality.

I now use graphs to illustrate the analysis starting with the case in which $p = p^*$. When $p = p^* = 0$, $U_q^F = U_q^*$ but $U_s^F \neq U_s^*$ and so the slope of $U_q = 0$ under fixed fee for service is identical to that under the efficient case. As already explained, $s^F < s^*$ for any given q and so U_s^F is below U_s^*. The first best solution then is E^*, shown in *Figure 6* and thus making $\hat{E}$ a second-best solution with $q^F > q^*$ and $s^F < s^*$ under Cases I and $q^F < q^*$ and $s^F < s^*$ under Cases II.

Figure 2.6: Fixed Fee for Service and Efficiency

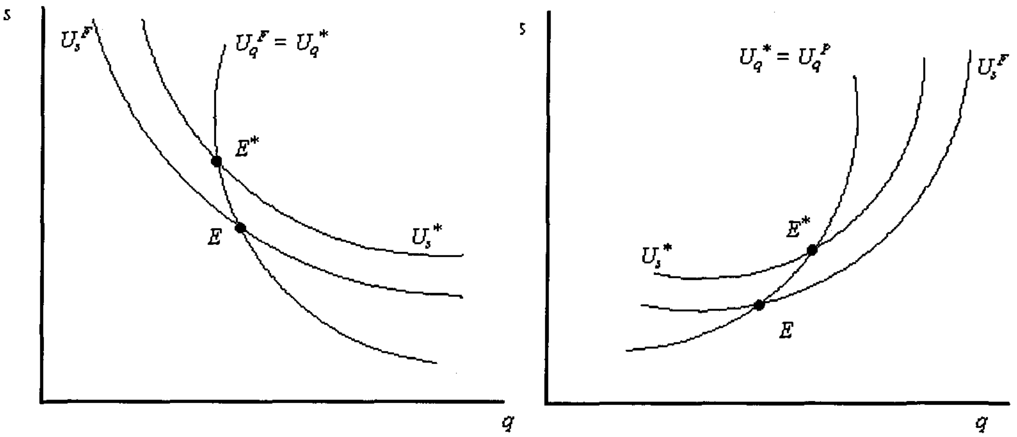


Figure 6a *Figure 6b*

PROPOSITION 6: *The equilibrium under the fixed fee is second best, even if the reimburser uses the efficient fee, with more (less) quantity than the first-best quantity under Cases I (Cases II). The second-best quality is less than that of the first-best regardless of the Cases but welfare is higher under Cases I than Cases II.*

The reason for the inability to achieve efficiency in both quality and quantity is the non-contractibility of quality. The reimburser has only one instrument, the fee, to induce two behaviors. When the fee is used to induce the efficient supply of quantity, it cannot be used again to induce the efficient supply of quality unless the physician is perfectly altruistic.

These results are different from Ma (1998), where quality is contractible and so the fixed fee induces the socially desirable level of quality. In the current paper when the fee achieves the efficient level of quantity, low supply of quality results. Chalkley & Malcomson (1995) show the physician can be induced to provide the efficient level of quality as long as the fee is sufficiently high to cover cost. Similarly, Kesteloot and Voet show that quality increases when the fee increases. The results in the current paper show that increasing the fee (above p^*) can increase quality only under Cases II but decreases it under Cases I. However setting the fee above p^* implies that neither **(8)** nor **(9)** is satisfied and so the resulting equilibrium has a lower level of welfare. Ma as well as Chalkley & Malcomson does not specify quantity in their model and quality is contractible, and so their model is not able to capture the inability of fee for service to achieve efficient quantity and quality together. Kesteloot and Voet include quantity but focus only on the case in which quality and quantity are complements.

2.4.2 Prospective Payment

Under the prospective payment, **(1)** and **(2)** show that the efficient supply of quantity and quality are not achievable regardless of α. The reimbursement, G, does not affect q^P and s^P and so the reimburser cannot use any reimbursement tool to induce efficient supply. Equation **(1)** shows that for any given s, $q^P < q^*$. *Section 3.2* shows that $q^P < q^F$ and $s^P > s^F$ under Cases I and $q^P < q^F$ and $s^P < s^F$ under Cases II. *Section 4.1* also shows that when p is low ($p < p^*$) $q^F < q^*$. It follows that $q^P < q^F < q^*$ regardless of the category of cases. In the case of quality, *Section 3.3* shows that $s^P > s^C$ under Cases I. *Section 4.3* will show that when τ is low ($\tau < \tau^*$), $s^C > s^*$. It follows that $s^P > s^*$ under Cases I. Similarly, under Cases II, *Section 3.2* shows that $s^P < s^F$ and *Section 4.1* shows that $s^F < s^*$. It follows that $s^P < s^F < s^*$ under Cases II. Thus while too little quality is supplied under the fixed fee for service, prospective payment, under Cases I, induces a higher level of quality than the first-best. If the objective of the reimburser is to get a high level of quality then it can choose prospective payment for Cases I. However, the high quality under Cases I, comes at the cost of much too little quantity compared to fixed fee for service. Note that the first-order conditions under prospective payment both deviate from **(8)** and **(9)** and so the resulting equilibrium has a lower level of welfare than that under the fixed fee.

I now use graphs to illustrate the analysis. For any given s $q^F < q^*$ and for any given q, $s^F < s^*$ and so U_q^P (respectively, U_s^P) lies to the left of U_q^*, (respectively, U_s^*) and the resulting equilibrium is shown as P in *Figure 7* where $q^P < q^*$ and $s^P > s^*$ under Cases I and $q^P < q^*$ and $s^P < s^*$ under Cases II.[49]

Figure 2.7: Efficiency and Prospective Payment

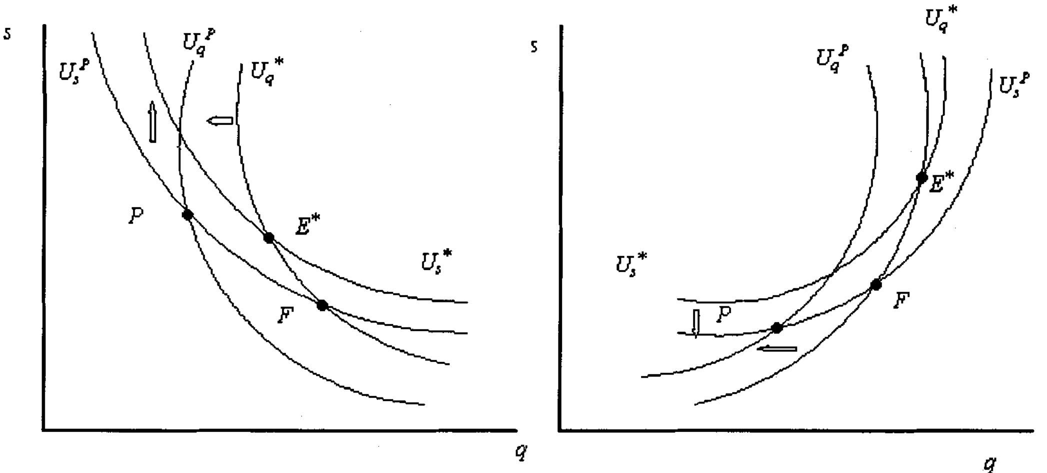


Figure 7a *Figure 7b*

PROPOSITION 7: *Efficient quantity and quality cannot be achieved under prospective payment regardless of the degree of altruism. The level of quantity and quality are too little under Cases II but quantity is too little and quality too much under Cases I.*

The results here are contrary to Ellis and McGuire (1986) where only the perfectly altruistic physician supplies the efficient quantity. The results here show even the perfectly altruistic physician supplies too little quantity. The reason for such difference comes from the reimburser's objective functions. In Ellis and McGuire, the reimburser cares about the patient and the cost of resources and so does not take the physician's welfare into account. The results in the current paper then implies that when the physician's utility is taken into account, then her degree of altruism becomes less important, under prospective payment, in determining the level of quantity and quality of care.

The result of too much provision of quality under Cases I is contrary to Ma and McGuire (1997), where the prospective payment equilibrium is characterized by too little quality and too much quantity when quantity and quality are complements or substitutes and the patient's

[49] With $(1+\alpha)B_q(q,s) - c_q(q,s) = 0$ then for any given s, $B_q < c_q$ under the efficiency case. Under the prospective payment $\alpha B_q(q,s) - c_q(q,s) = 0$ implies $B_q > c_q$ for any given s. The concavity assumption implies that $q^F > q^*$ for any given s. A similar reasoning shows that $s^F < s^*$ for any given q.

marginal benefit of quality is high relative to that of quantity. As already explained, Ma and McGuire do not take cost cross effect into account. The inclusion of cost cross effect in the current paper has made it possible to show that high levels of quality are possible under prospective payment. The physician supplies too little quantity but has the incentive to supply more quality under Cases I than under Cases II.

The physician has the incentive to supply a high level of quality when quantity and quality are complements and the cost cross effect is positive and strong as under Cases I. Because low supply of quantity implies low marginal cost of quality when the cost cross effect is positive, quality increase in this case is consistent with profit maximization. A strong positive cost cross effect, then, provides the incentive for the physician to increase the quality effort per visit or hospital day in order to reduce the number of visits or length of stay. For example, performing a surgery (quality effort) on a patient requires in-hospital stay (quantity) and so quality and quantity are complements. If the quality effort is skill-intensive and so leads to an increase in the marginal cost of hospital stay per day, then the physician, because she bears the full cost of quantity under prospective payment, has the incentive to increase the quality effort per day in order to reduce the patient's length of stay at the hospital.

2.4.3 Cost Sharing

The relationship between the equilibrium quality and quantity under cost sharing and the efficient case depends on the sizes of τ and α. Equations **(1)** and **(2)** are identical to **(8)** and **(9)** when $\tau = \tau^* = \frac{B_q}{c_q} = \frac{B_s}{c_s} = \frac{1}{(1+\alpha)}$. Thus when $\tau = \tau^*$, $q^C = q^*$ and $s^C = s^*$. This is different from the $\tau^* = (1-\alpha)$, in Ellis & McGuire (1986) (for quantity). The reimburser in the current model is able to use one policy instrument, τ, to induce the optimal provision of both quality and quantity even if it does not know the degree of altruism. This is made possible because, under cost sharing, the reimburser covers part of the cost, which is a function of quantity as well as quality. This makes the non-contractibility of quality less consequential than in the case of fixed fee. Thus cost sharing is the most effective in inducing the efficient supply of quality without compromising the efficiency of quantity. Efficient levels of quality and quantity can be induced by estimating the marginal benefit and marginal cost of quality or quantity. Thus cost sharing is the best payment scheme if the reimburser knows the degree of altruism and wants to induce efficient supply of both quantity and quality.

The α can be used to compute the range for τ^*. When the physician does not care about the patient and only cares about profit, ($\alpha = 0$), she will only produce the minimum quantity and quality unless the payment scheme is such that she does not bear the cost of treatment, $\tau = 1$. Under full cost reimbursement ($\tau = 1$), cost saving or cost increase does not influence the physician's decision when choosing the level of quality. Thus setting $\tau^* = 1$ induces the physician to increase the supply of quality and quantity towards the efficient levels. When the physician is perfectly altruistic, ($\alpha = 1$), $\tau^* = ½$. This is different from Ellis and McGuire where $\tau^* = 0$ implying prospective payment, when the physician is perfectly altruistic. The current paper shows that the reimburser needs a tool to induce the desired supply even when the physician is perfectly altruistic. When the physician is less altruistic, setting $½ < \tau^* < 1$ makes the reimburser able to create the incentive for quality and quantity improvement at the desirable levels.

Graphically, when $\tau = \tau^*$ then $U_q^* = U_q^C$ and $U_q^* = U_q^C$. The resulting equilibrium has $q^C = q^*$ and $s^C = s^*$ as shown as E^* in *Figure 8*. When $\tau > \tau^*$, then both U_q^C and U_s^C are to the right of U_q^* and U_s^* respectively. Indeed, given $B_{qq} < 0$ and $c_{qq} > 0$ as well as $B_{ss} < 0$ and $c_{ss} > 0$, then for a given s (respectively q) the physician supplies more q (respectively s) under cost sharing than is efficient. *Figure 8* shows that the resulting equilibrium, C, has $q^C > q^*$ and $s^C < s^*$ under Cases I and $q^C > q^*$ and $s^C > s^*$ under Cases II as confirmed by (6) and (7). Similarly, when $\tau < \tau^*$then $q^C < q^*$ and $s^C > s^*$. This is not shown on the graph.

Figure 2.8: Efficiency and Cost Sharing

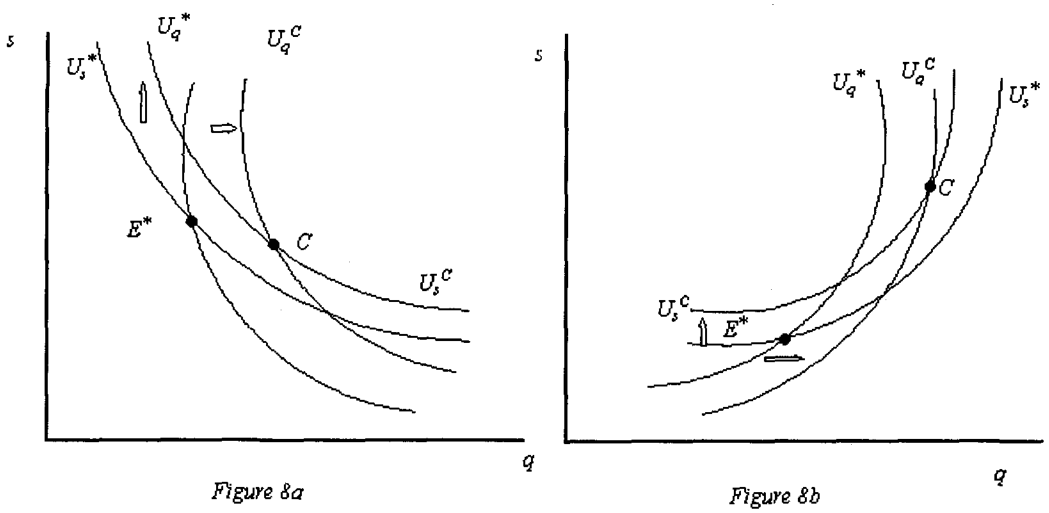


Figure 8a *Figure 8b*

PROPOSITION 8: With cost sharing, $\tau = \frac{B_q}{c_q} = \frac{B_s}{c_s} = \frac{1}{(1+\alpha)}$ ensures that the efficient levels of both quantity and quality are chosen. Thus, the reimburser need not know the degree of altruism. When $\tau > (<) \frac{B_q}{c_q} = \frac{B_s}{c_s} = \frac{1}{(1+\alpha)}$ then quantity is over (under) supplied in both categories of cases but quality is under (over) supplied under Cases I and over (under) supplied under Cases II.

These results are different from Ma and McGuire (1997), where efficiency under cost sharing is only possible when quality and quantity are substitutes. The results in the current paper show that the efficient level of cost sharing does not change with the relationship between quality and quantity in the patient's utility or the cost cross effect. Consequently, from efficiency perspective, the results rank cost sharing above the other two payment schemes because it is the only payment scheme that can achieve the first-best equilibrium. However, the ability of cost sharing to induce the efficient levels of quantity and quality depends on the knowledge of the degree of altruism, which may not be observable. Even though fixed fee for service provides a higher level of welfare than the prospective payment, and also requires the knowledge of the patient's marginal benefit it can only induce a second-best equilibrium. The results also ranks prospective payment scheme the lowest since it provides the lowest level of welfare.

2.5 Conclusion

This paper has shown that the relationship between quality and quantity in both the patient's utility and the cost of care are crucial in determining the equilibrium quality and quantity that the physician supplies under different payment schemes. Combining the two relationships helps identify the conditions under which financial incentives induce high levels of quality and the conditions under which it does not. When quantity and quality are positively related then increasing the provider's payment is an incentive for the provision of high levels of quality. When quality and quantity are negatively related however any increase in payment reduces quality enhancement effort. Hence prospective payment, because it provides the lowest payment at the margin, induces the lowest level of quality when quantity and quality are positively related but induces the highest level of quality when quantity and quality are negatively related.

The paper has also shown that neither fixed fee for service nor prospective payment can achieve efficient provision of quantity or quality of care even when the physician is perfectly altruistic. Cost sharing on the other hand can achieve the efficient levels of quantity and quality as long as the reimburser knows these levels. However, the model shows that the efficient levels of quantity and quality can only be computed if the reimburser knows the level of the physician's altruism.

There is no obvious direct way to compute the level of the physician's altruism. A good indicator of the degree of the physician's altruism could be his behavior towards uninsured patients such as illegal immigrants, visitors and international students. Because such patients are not likely to be able to pay for services they receive, a physician that treats them has to be prepared to offer free services. The proportion of such patients that are treated relative to the total that shows up is then a good indirect measure of the degree of altruism. Alternatively, the reimburser could compare the intensity of treatment that an uninsured patient receives relative that which an insured patient with similar illness receives. Of course the physicians should not know these because if they did, severe moral hazards problems may result.

The paper shows that when the marginal benefit of quantity is high relative to marginal cost then efficiency demands that the reimburser increases the cost coverage under cost sharing. This implies that the cost coverage for treatments that improve patient's health should be higher than those that do not improve health. Alternatively, efficiency demands that the reimburser covers a greater percentage of cost when technological improvement reduces the marginal cost of care. To achieve efficiency with cost sharing the reimburser should cover at least half of the cost of care.

2.6 Reference List

Chalkley M. & J.M. Malcomson (1995), "Contracting for Health Services with Unmonitored Quality," *Discussion Paper*, Department of Economics, University of Southampton.

Ellis, R.P. & T.G. McGuire (1986), "Provider Behaviour Under Prospective Reimbursement, Cost Sharing and Supply," *Journal of Health Economics*, 5, 129-51.

Hanchak, N., N. Schlackman, & S. Harmon-Weiss (1996), "U.S. Healthcare's Quality-based Compensation Model," *Health Care Financing Review,* 17, 143-59.

Kahn, K., W. Rogers, L. Rubenstein, M. Sherwood, E. Reinisch, E. Keeler, D. Draper, J. Kosecoff, and R. Brook. (1990), "Comparing Outcomes of Care Before and After Implementation of the DRG-Based Prospective Payment System," *Journal of the American Medical Association*, 264(15), 1984-73.

Keeler, E. B. (1990), "What Proportion of Hospital Cost Differences is Justifiable?," *Journal of Health Economics*, 9(3), 359-65.

Kesteloot, K. & N. Voet (1998), "Incentives For Cooperation in Quality Improvement Among Hospitals – the Impact of the Reimbursement System," *Journal of Health Economics*, 17, 701-28.

Kosecoff, J., K. Kanh, W. Rogers, E. Reinisch, M. Sherwood, L. Rubenstein, D. Draper, C. Roth, C. Chew and R. Brook (1990), "Prospective Payment System and Impairment at Discharge: 'The Quicker and Sicker' Story Revisited," *Journal of the American Medical Association,* 264(15), 1980-83.

Laffont, J. J. and Tirole, J. (1989), "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94(3, Part 1), 614-41.

Ma, C.A (1998), "Cost and Quality Incentives in Health Care: Altruistic Providers," *Discussion Paper,* Department of Economics, Boston University.

Ma, C.A. & T.G. McGuire (1997), "Optimal Health Insurance and Provider Payment," *American Economic Review*, 87(4), 685-704.

Norton, E. (1992), "Incentive Regulation of nursing Homes," *Journal of Health Economics*, 11, 105-28

Nyman, J. (1985), "Prospective and 'Cost-plus' Medicaid Reimbursement, excess Medicaid Demand, and the Quality of Nursing Home Care," *Journal of Health Economics,* 4, 237-59.

Pope, G. (1990), "Hospital Nonprice Competition and Medicare Reimbursement Policy," *Journal of Health Economics*, 8, 147-72.

Rickman, N. and A. McGuire (1999), "Regulating Provider's Reimbursement in a Mixed Market for Health Care," *Scottish Journal of Political Economy*, 46(1), 53-71.

Schleifer, A.(1985), "A Theory of Yardstick Competition," *The RAND Journal of Economics,* 16(3), 319-27.

Wyszewianski, L., J. Thomas, & B. Friedman (1987), "Case-based Payment and the Control of Quality and Efficiency in Hospitals," *Inquiry*, 24, 17-25.

CHAPTER 3: THE EFFECT OF HOSPITAL DOWNSIZING IN BRITISH COLUMBIA ON THE QUALITY OF CARE FOR MATERNITY PATIENTS

3.1 Introduction

From the mid 1980s to 1993, the hospitals in British Columbia (BC) faced limited growth in their budgets because of economic recession, government deficits and the high cost of borrowing, (McGrail et. al., 2001). This was reflected in the gradual reduction in hospital inpatient use and capacity (McGrail et. al., 2001). However, inpatient use dropped sharply in 1994, when the Closer to Home Fund was established to support community or public care and so reduce hospital care (McGrail et. al., 2001; Hansard, 1995).

In the case of maternity care, the fund allowed hospitals to send low risk mothers and newborns home soon after childbirth so that they could receive care at home through community care providers. (Hansard, 1995; BC Reproductive Care Program, 2002). Since all medical care and hospitals are publicly funded, the cost of care to the government is likely to be lower when the care is delivered at home than when it is delivered in the hospital. The objective of the policy-induced reduction in length of hospital stay then was to reduce the cost of care to government, (Hansard, 1995). However, this transfers the cost to patients, their families and friends.

Concerns have been raised about the consequences of earlier discharge on the health of maternity patients (BC Reproductive Care Program, 2002). However, no study has yet been done on the effect of the policy on health outcomes of maternity patients. This paper estimates the effect of the hospital downsizing policy on the readmission rate of maternity patients. A change in readmissions could represent a change in rates of morbidity, and therefore can be interpreted as an indicator of changes in health outcomes. Investigating the effect of the policy on the readmission rate also contributes to our understanding of the actual magnitude of cost savings. Because readmissions are expensive, a small increase in readmissions could substantially offset the cost savings from the early discharge policy (Weissman et al., 1994). The effect of the downsizing on readmission rates therefore provides necessary information for assessing the cost effectiveness of the policy.

To my knowledge, the studies that use maternity data to examine the effect of the reduction in length of stay use the United States (US) data. Maternity care in the US hospitals is funded through both private insurance and public insurance (under the Medicaid Program) which is administered by private companies such as the Health Maintenance Organization (HMO). While the other private insurance are likely to reimburse care by the fee for service scheme, the HMOs use prospective payment or capitation systems. Prospective payment refers to a payment scheme under which the reimburser makes a fixed payment to the health care provider regardless of the total quantity and the quality of care. Prospective payment is similar to downsizing in that both force health care providers to adopt cost control policies such as reduction in length of stay.

Gazmararian and Koplan (1996) compare the length of stay after delivery and readmission rates of mothers under different types of insurance plans. They find that length of stay varies with insurance plan, and is lower for HMOs. This result confirms the perception that the incentive to minimize cost under prospective payment comes at the expense of reduction of quantity of care (length of stay). However, the authors of this study do not find an association between the readmission rate of maternity patients and length of stay or plan type.

Tai-Seale et al., (2001), use longitudinal data on Medicaid patients in three counties in California to examine the effect of capitation on the use of obstetric services. Like prospective payment, under capitation the health care provider receives a fixed payment per patient regardless of how much care is provided. They compare length of stay and readmission rates of patients under fee for service to those under capitation. Like Gazmararian and Koplan (1996), they find that the cost saving that accompanies capitation comes at the expense of reduction in the provision of prenatal care and delivery length of stay but does not cause any significant change in readmission rates.

Several other papers use non-maternity data to examine the effect of prospective payment and its accompanying reduction in length of stay on readmission or mortality rates. The results of these papers are mixed. Some authors find that after the introduction of the prospective payment patients are more likely to be discharged in an unstable condition (Kosecoff, et. al, 1990; Rubenstein, et. al., 1990) and to be ill at the time of admission (Keeler, et. al., 1990), and that readmission and mortality rates increased (Keeler et al 1990). Other authors find no significant effect on readmission or mortality rates (Manton et al. 1993, DesHarnais et al. 1987) while Kahn et al. (1990) find that mortality rates fell.

This paper provides empirical estimates of the effects of the downsizing policy in BC on the length of stay and readmission rates of maternity patients. The results show that the length of stay following delivery decreased, likely reducing the public cost of maternity care. Maternal readmission rates increased, especially for mothers who did not experience medical complications (the low risk mothers) and for aboriginal mothers.

The paper is organized as follows. Section 2 describes the data. Section 3 focuses on the estimation and results of the length of stay equation. Section 4 estimates and reports the results of the readmission equation and Section 5 concludes.

3.2 Data Analysis

The paper uses maternity data on all deliveries in seventeen acute care hospitals in BC from fiscal year 1993 through fiscal year 1996. The data, which are administered and provided by the BC Ministry of Health and Ministry Responsible for Seniors, consist of hospital records that include information on a host of variables including age, procedure, diagnosis, hospital, local health areas, dates of admission and discharge, patient's aboriginal status, doctor's specialty as well as transfers between hospitals. The seventeen hospitals included in the sample, which form about 21% of the eighty hospitals that provide obstetrics in the province, were chosen to provide a broad geographical representation of the province.

Of the seventeen hospitals, four are located in the interior region of the province, three on the northern part of the province, three on the Vancouver Island and the remaining seven in the Lower Mainland, with four in the Vancouver and Richmond area, and three in the Fraser valley. Selecting about half of the hospitals from the Lower Mainland is consistent with the population distribution in the province because about half the population of the province is concentrated in the Lower Mainland. During the four-year period of study, 92,594 deliveries and 3939 maternity readmissions occurred in the selected hospitals. This comprises about 50% of the total maternity cases in the province during the period (B.C Vital Statistics, 2001).

This study isolates aboriginal mothers from non-aboriginal mothers because of concerns about aboriginals' health. Aboriginal people in BC are more likely than non-aboriginals to live in rural areas (Shrier & Ip, 1994). In addition, aboriginal women are more likely to be teen parents, single mothers, low-income earners, and victims of physical and substance abuse (Health Canada, 2002). Aboriginals are also likely to face barriers accessing health care because of their language and culture (Health Canada, 2002). Consequently aboriginal women on average have poorer health outcomes than non-aboriginals (Health Canada, 2002). It is therefore of considerable policy interest to explore any differences in the effect of the policy change on health outcomes of aboriginals and non-aboriginals. Obviously non-aboriginals is a cluster of many races and so it would have been more interesting to include other races, but such information was not available.

Table 1a shows that the patients have an average age of 29.4 and about 4.3 of them were readmitted.[50] While 20.6% of them underwent Caesareans, more than 65% of them had complications.[51] About 3.2% of the patients are aboriginals who are on average five years younger than the non-aboriginals and are less likely to undergo Caesareans or have complications but are more likely to be readmitted. Though aboriginals form 3.2% of the deliveries they form 5% of the readmissions.

The high percentage of patients with complications shows how common such complications are among women in maternity. For example, because pregnancy increases pressure on the kidneys and the bladder, it is common for a woman's blood pressure to change or for her to have circulation problems during pregnancy. In addition, lactation problems can be common. However, a lot of these complications are mild and cease after childbirth. Some complications however develop or persist after discharge and so cause readmissions. Information on the severity of these complications would show the degree of severity that can cause readmission. This information is however not available. Because patients without complications are likely to be low risk and the policy is likely to send low risk patients home early, it is interesting to control for complications and study how the effect of the policy varies according to complications.

Table 1b shows the trend of the proportion of patients who had complications and those who underwent Caesareans over the four years of study. The percentage of patients who underwent Caesareans or had complications remained fairly stable throughout the four years of study for the whole sample and for non-aboriginals as well. Thus, in general there was no significant change in the proportions of the patients who had complications or who underwent Caesareans.

Figure 1 shows the frequency distribution for length of stay for each year. The graph shows that the proportion of patients who were discharged after one to three days increased over

---

[50] All Tables and Figures for this chapter can be found in Appendix B.

[51] Complications refer to haemorrhage of pregnancy, complications in labour and delivery as well as complications of the puerperium. Complications of pregnancy include hypertension, diabetes and anaemia developed during pregnancy. Complications of labour and delivery include postpartum haemorrhage and damage to pelvic joints and ligaments. Examples of complications of the puerperium include infection of the nipple and failure of lactation.

the four years. Thus, the proportion of those who stayed for four or more days decreased over the four years. While the distribution for 1993 peaked at three days, the rest of the years had their peaks at two days. The graph also shows that frequencies for the last two years are almost identical implying a similar pattern of length of stay in each of the two years after the policy change.

The first panel of Table 2 shows that, with the exception of patients with complications whose length of stay increased in 1994, length of stay decreased steadily during the four years of study. In general, as shown in the last column of Table 2, patients who underwent Caesareans stayed longer than those who had vaginal delivery and patients with complications stayed longer than those without complications. There is a greater difference between the length of stay of patients who had Caesarean and those who had vaginal delivery than between those who did or did not have complications. Length of stay decreased more for patients who underwent Caesareans than for those who had vaginal deliveries. As shown in the last two panels of Table 2, with the exception of aboriginals with complications and Caesareans, the steady reduction in length of stay occurred for both aboriginals and non-aboriginals but was greater for non-aboriginals. The policy must have increased length of stay for at least some of the most vulnerable mothers, aboriginals with complications and/or underwent Caesareans. In general however the policy was successful in reducing the length of stay for maternity patients and so is likely to have reduced the hospital cost of delivery.

Figure 2a and 2b show the cumulative readmission rates for the first 7 and 90 days after discharge. The readmission rates for 1994 exceed those of 1993 throughout the first 90 days. As shown in Figure 2a, 1994 had the highest readmission rate at the end of the 90 days. Since 1994 is the transitional year this increase in the readmission rate may at least be partly due to transitory detrimental effect of the policy on patients.

An increase in the readmission rate in the period immediately after discharge indicates a deterioration in health outcomes that could be prevented through a longer hospital stay (Weinberger et al., 1988). When patients are sent home too early and in unstable condition, it is more likely that they will be readmitted shortly after discharge rather than later because health stability increases with time. In general, most readmissions during the first seven days after discharge reflect the existence of premature or sub-optimal discharge (Welch, et al., 1992). This may explain why daily readmission rates are highest soon after discharge, reflected in the steep slope of the readmission rates plot close to the origin. Figure 2b shows that 1996, when the average length of stay was shortest, had the highest readmission rate in the early days following discharge.

While the 7-day readmission rates reflect the effect of length of stay on patient stability immediately following discharge, the 90-day readmission rate also includes any longer-term effects of the policy on patient health. Examples of diagnoses that cause readmission in the later days after discharge include haemorrhoids, psychoses, as well as types of complications mentioned above. Ninety days is the approximate length of time it takes for a woman to make a full physical recovery, which takes up to sixty days, and adjust emotionally following childbirth, which takes at least ninety days (BC Reproductive Care Program, 2002). Any effect of the policy on women's health outcomes therefore will be realized within these ninety days.

Table 3a and 3b show the 7-day and 90-day readmission rates respectively when patients are classified according to race, the type of delivery and the occurrence of complications. In general the readmission rates of aboriginals exceed those of non-aboriginals. Both Tables show that regardless of race and year, the readmission rates for patients who underwent Caesareans exceed those who had vaginal delivery. For non-aboriginals the readmission rates of patients with complications slightly exceed those without complications. With the exception of the 7-day readmission rates for vaginal deliveries for aboriginals, which are fairly constant, the readmission rates for the other categories of patients fluctuated over the years. Because non-aboriginals dominate the sample the fluctuations of non-aboriginals are similar to those of aboriginals and non-aboriginals combined. The fluctuation of the readmission rates for aboriginals however differs from that of non-aboriginals. Thus, in general the health outcome for aboriginals is different from those of non-aboriginals and so isolating aboriginals helps reveal the difference.

In summary, the tables so far show that length of stay decreased and readmission rates increased in the later years of the sample. The increase in readmission rates could be an indication that the home care made possible by the Closer to Home Fund probably was either not well organized and so some patients did not receive care on time or that the quality of care that the patients received at home was lower than what they received in the hospital.[52] However, these changes cannot readily be attributed to the policy change because they could be due to changes in patients' characteristics. Thus to find out the extent to which the policy caused length of stay and readmission rates to change, holding patients' characteristics constant, I use regression analysis to estimate the effect the downsizing on length of stay and readmission rates.

---

[52] Note that the home care program for maternity patients is only one of the variety of home care programs supported by the Closer to Home Fund. The fund provided home care for cancer patients as well.

3.3 The Length of Stay Equation

3.3.1 Empirical Framework

The number of days a maternity patient spends in hospital following delivery will depend on the type of delivery (vaginal or Caesarean section), the patient's characteristics that affect the rate of healing, and on the policy in effect at the time of delivery. Of the relevant patient characteristics, data are available on the patient's age, race and whether she experienced health complications. Policy changes are captured through dummy variables representing years in the sample period. The specification of the regression model is as follows:

$$LOS_i = \beta_0 + \beta_1 DYEAR_i + \beta_i Z_i + \beta_j DYEAR_i * Z_i + u_i \qquad \textbf{(1)}$$

where LOS_i is the length of stay in days of patient i in the hospital, $DYEAR$ is a vector of dummies for the year of delivery, and Z_i includes an indicator of the method of delivery as well as the patient's, aboriginal status, an indicator of whether she has experienced health complications, and her age, and β_j is a vector of corresponding coefficients. The variables in Z_i are also interacted with the year of delivery dummies to allow for differential effects of the policy on the length of stay of patients with different characteristics. A dummy variable for aboriginal status is included because of the well-known differences between the average health conditions and outcomes of aboriginal and non-aboriginal people. Complications are coded according to the "most responsible diagnosis." If this does not fall under ICD-9 Codes 650 – 659.99, which refer to normal delivery, and care during pregnancy, labour, delivery, a patient is coded as having experienced complications. As already mentioned, diagnoses that are considered as complications include haemorrhage of pregnancy as well as complications in labour and delivery. Age is included because it is an important factor in maternal health and so affects how a patient responds to care. To ensure flexibility of the relationship between age and length of stay I include the square and inverse of age. The model is estimated using ordinary least squares.

3.3.2 Results

The estimated coefficients of the length of stay equation are reported in Table 4. The first column of Table 4 shows the results of the length of stay equation without interactions with the year dummies.[53] The results show that after controlling for age, complications and type of delivery, aboriginals on average stay longer than non-aboriginals. The slope of the length of stay equation with respect to age is – 0.134 +0.004*age which is zero when age is 33.5 implying that the effect of age on length of stay divides the patients into two depending on whether they are younger or older than 33.5. Within the young group, younger mothers stay longer than older mothers. This is perhaps because younger mothers, who are likely to be teenagers and in early twenties, require more assistance in taking care of themselves than do older mothers who may have experience of childbirth. However, in the older group older mothers are likely to stay longer than younger mothers. This is probably because, given that complications are controlled for, mothers above thirty-three are likely to respond slowly to care and so require more care or monitoring.

There is no statistically significant difference in the length of stay for patients with complications and those without complications. This is not surprising because, as already explained, many of the cases classified as complications may be mild. However, the length of stay of patients with Caesarean deliveries is greater than those with vaginal deliveries. The year dummies show that length of stay decreased steadily through the four years. The reduction in length of stay is highest in 1995 and lowest in 1996. Thus, even though the policy continued to reduce length of stay over the years the reduction decreased after the second year of the policy.

The second column of Table 4 shows the results of the length of stay equation when the year dummies are interacted with the other variables. The results show no statistically significant difference in the reduction in length of stay for the different categories of patients. Thus, the variations in the reduction in lengths of stay among the different categories of patients observed in Table 2 are not statistically significant.

The policy was effective in reducing the length of stay of patients and so must have succeeded in reducing the hospital cost of care. It however transferred cost partly onto public care through the home care program and partly to the patient and their families as they take care of the mothers. Since the reduction in hospital care is replaced by home care, the patient is made worse off if the quality of the home care is below that of the hospital. To find out how this policy affected the quality of care I now examine the extent to which it increased the rate of readmissions.

---

[53] Since length of stay is non-negative, functional form estimation such as Poison, hazard or logit should typically estimate it. However, when the length of stay equation was estimated by a hazard function, the results were not qualitatively different from those of ordinary least square estimation and so the analysis focuses on the results from ordinary least square estimation.

3.4 The Readmission Equation

3.4.1 Empirical Framework

Since quality of care refers to the effectiveness of treatment, there is a negative relationship between quality of care and the readmission rate. The production function for quality of care can be written as Q = *F(hospital effort input, home effort input)*. Quality of care is positively related to the effort input by the health care provider. Effort input refers to all service inputs by the hospital that contribute to the outcome of a given stay in the hospital. Following Ma and McGuire (1997), I define effort broadly to include all effort inputs that improve patients' comfort as well as those that improve treatment, such as monitoring the patient's condition to match a given problem with a specific therapy. The home effort input refers to the care that the patient receives at home from family and friends. The production function for the readmission rate then can be written as: R = *F(Q, patient characteristics)* which in turn implies that R = *F(hospital effort input, home effort input , patient characteristics)*. I use type of delivery (Caesarean or vaginal) to represent effort input. I use the same vector of characteristics that were included in the length of stay equation: age, complications, and race. Again, these characteristics are interacted with the year dummies.

Social factors such as education, marital status (or support from friends and family), and income may also be important in determining the readmission rate. For example a married mother or a woman who does not live alone may be more likely to receive help at home (home effort input) and so be less likely for her condition to worsen after discharge than a single mother who lives alone. Even if she lives alone, a high-income mother may be more likely to afford hiring a nanny and so may receive a better care at home after discharge than a low-income mother and so be less likely to be readmitted. Finally a mother with high education may be more likely to take care of herself at home than one with low education and so the higher the education of the mother the less likely she may be readmitted.

These variables are not included in the specification above simply because they were not available. The omission of such variables in the estimation can bias the results if the omitted variables are correlated with the variables in the equation. The most likely variable in the estimation to be correlated with these social factors is aboriginal status. Having home support, income and education may not affect the type of delivery, age or complications. Reasons for Caesareans include mother's bone structure, fetus' health, as well as mother's age and obstetrics health. The complications in the readmission equation refer to complications before discharge and they may not be correlated with income, marital status, or education. As already explained, aboriginals are likely to be low-income earners and single mothers. Thus the omission of the social factors in the readmission equation is likely to affect results on aboriginals and so the results are interpreted with caution.

3.4.2 Method of Estimation

I use the discrete duration model as described in Kennedy (1998) to find the effect of the change the independent variables on the readmission rate. A duration model, rather than logit or probit models that do not take time into account, is more appropriate in assessing the effect of the downsizing on the readmission rate. As already explained, changes in readmission rates shortly after discharge reflect changes in length of stay. A duration model described below is able to factor in the effect of time on the probability of readmission. I now describe the duration model for the specification.[54]

Let $f(t)$ be the probability of being readmitted at time t after discharge and F(t) be the cumulative probability of readmission by time t. The survival function, the probability of not being readmitted to hospital during the first t days following discharge, is defined as $S(t) = 1 - F(t)$. The hazard function, $\lambda(t)$, is the probability of being readmitted t days after discharge, conditional on not being readmitted previously. I use the estimates of the hazard function to compute the probability of interest, the unconditional probability of being readmitted t days after discharge. The relationships between the hazard function, the unconditional probability and the survival function are as follows:

$$f(t) = \lambda(t)S(t), \qquad \lambda(t) = -\frac{d\ln S(t)}{dt} \tag{2}$$

It follows from the above that: $S(t) = \exp[-\int_0^t \lambda((u)du]$. This relationship allows the likelihood function to be written in terms of the hazard function. When the patient is readmitted during the sample window, her observation enters the likelihood function as $f(t)$ but when the patient is not readmitted the observation enters as $S(t)$. The likelihood function for the continuous time model is:

[54] See Kennedy (1998) and Greene, (1993)

$$L = \prod_{i=1}^{N} \lambda(t)^{\delta_i} \exp[-\int_0^t \lambda(u)du] \quad (3)$$

where N represents the number of observations and δ_i equals one when the patient is readmitted at time t and equals zero otherwise[55]. The patients that are not readmitted after 90 days are the censored observations and those who are readmitted are uncensored. I estimate the model over the first 90 days following the initial discharge.

The data provide the dates but not the time of day of discharge and readmission , and so are discrete. With discrete-time estimation, a likelihood function is built for each of the 90 days after discharge. For the first day after discharge a patient is either readmitted or not readmitted. The likelihood function for the day captures this. For each patient, whether the she is readmitted or not determines her contribution to the day 1 likelihood function. For the second day those who have not yet been readmitted are either readmitted or not readmitted and so the likelihood function for this day also captures such difference. Thus patients contribute several observations to the likelihood function. For example, a patient that is readmitted on the fourth day after discharge contributes four observations, one each for day one to three where she is not readmitted and one on the fourth day when she is readmitted. She does not appear in the likelihood functions for the remaining eighty-six days. Thus the censored individuals appear in all the likelihood functions for all the ninety days and so each would contribute ninety observations. The product of the ninety likelihood functions for the individual days produces the likelihood function for the full sample. I specify a logit functional form for the hazard model. The likelihood function therefore can be estimated using a standard logit procedure.

To choose a baseline hazard for the estimation, I compute the Kaplan Meier estimates and graph them in Figure 3.[56] The pattern of the estimates shows a downward trend over time with a spike on the second day. In addition the estimates decrease at a decreasing rate and so reveal a convex pattern. Adding a baseline to the characteristics already included in the model yields the following specification for the logit:

---

[55] Note that $f(t)S(t) = [\lambda(t)S(t)]^{\delta i} S(t)^{1-\delta i} = \lambda(t)^{\delta i}S(t)$

[56] The Kaplan Meier estimates for the readmission rates show the fraction of the patients that are readmitted for the first time for each day. For example to get fraction for day five I divide the number of patients readmitted on the fifth day by all those that have survived in not being readmitted. Figure 3a shows that this fraction falls as the days after discharge increases. The Kaplan Meier estimates for the survival function (the opposite of the readmission) then shows the fraction of those that survive in not being readmitted. For each day I subtract those readmitted from those not yet readmitted and divide the result by those not yet readmitted. Figure 3b shows that this fraction increases as the days after readmission increases.

$$\Pr(y_{it} = 1) = \frac{e^{X_i'\beta + \gamma(t)}}{1 + e^{X_i'\beta + \gamma(t)}}$$

Where $X\beta = \beta_0 + \beta_1 DYEAR + \beta_i Z + B_j DYEAR * Z$

and

$\gamma(t) = \gamma_1 t + \gamma_3 t^2 + \gamma_2 T2$

The variable $T2$ takes on the value one when $t = 2$ and takes on the value zero otherwise. This dummy is included to incorporate the spike in day 2 in the Kaplan Meier estimates.

The likelihood function for the discrete estimation is:

$$L = \prod_{i=1}^{N_1} \frac{e^{X_i'\beta+\gamma(t)}}{1 + e^{X_i'\beta+\gamma(t)}} \prod_{j=1}^{N-N_1} \frac{1}{1 + e^{X_i'\beta+\gamma(t)}} \cdots \prod_{i=N_{89}}^{N_{90}} \frac{e^{X_i'\beta+\gamma(t)}}{1 + e^{X_i'\beta+\gamma(t)}} \prod_{j=1}^{N-N_{90}} \frac{1}{1 + e^{X_i'\beta+\gamma(t)}} \tag{4}$$

where X is a vector the explanatory variables, β is a vector of the coefficients to be estimated, N_1 is the number of those readmitted on the first day after discharge and N-N_1 is the number of those not readmitted on the first day. Thus a total of N-N_1 observations make it to the likelihood function for the second day. Out of these some are readmitted on the second day and so a total of N-N_2-N_1 observations make to the third day and so on. Thus for an individual, the probability of being readmitted on the tth day after discharge, having not been readmitted before t, is $\lambda(t) = \frac{e^{X_i'\beta+\gamma(t)}}{1 + e^{X_i'\beta+\gamma(t)}}$. The survival function at time t for an individual in (3) can be written as

$\prod_{t=1}^{T} \frac{1}{1 + e^{X_i'\beta+\gamma(t)}}$. For a censored individual, $T = 90$.

3.4.3 Results

The logit results are reported in Table 5. The first column of Table 5 shows the results of the readmission equation when no interactions between explanatory variables are included in the specification. The results show that after controlling for age, year and complications, aboriginals have a higher readmission hazard than non-aboriginals. As already explained the omission of social factors could bias the results on aboriginals. Thus the readmission hazard for aboriginals explains the extent to which the outcome of care is due to the genetic characteristics of aboriginals as well as the environmental, economic and social effects that are specific to the race.

The high readmission hazard for aboriginals is consistent with other evidence that aboriginals have poor health outcomes compared to non-aboriginals.

The coefficient on age is negative on the coefficient on its square is positive. Again the slope with respect to age is zero at 25.5 and so among the patients that are younger than 25.5 years, a younger mother has a higher readmission hazard than an old mother. As already discussed, a young mother may be less likely to take care of herself after discharge than an older mother. The readmission hazard however increases with age when the patient is older than 25.5 years. The readmission hazard for patients with complications is greater than that for those with no complications. This result is consistent with Table 3b that shows higher readmission rates for patients with complications when all the years are combined. The readmission hazards for the year dummies show a jump in 1994 a slight fall in 1995 and another increase in 1996.

The results also show that *CAESAREAN* is not statistically significant. This implies that after controlling for all the other variables, there is no statistically significant difference in the readmission hazards of patients who had caesarean deliveries and those who had vaginal deliveries. This differs from Table 3a and 3b where the readmission rates of patients who had Caesarean exceed those of patients who had vaginal delivery. The results in Tables 3a and 3b could be driven by age because older mothers are likely to undergo Caesarean. Thus given that age is controlled for the results from the regression then mean that the type of delivery per se does not affect the readmission hazard.

The results also show that time has a negative sign and is statistically significant. The T2 dummyis not statistically significant implying the spike at $t = 2$ in the Kaplan Meier estimate does not represent a statistically significant difference between $t = 2$ and the general trend of the readmission hazard with respect to time. The square of time has a positive but small coefficient and so the slope remains negative throughout the ninety days but becomes flatter with time revealing a convex relationship between the probability of readmission and time. Consistent with the Kaplan Meier estimates, these results imply that the readmission hazard falls as the days after readmission increase but at a decreasing rate.

The second column of Table 5 reports the results of the readmission equation with the interactions. The inverse of *AGE*, the interactions of *AGE* with the year dummies, of *T2* and the year dummies as well as the interactions of *CAESAREAN* with the year dummies were not significant and so were not included in the final specification reported in Table 5. The results show that the policy had a smaller effect on the readmission hazard of patients with complications and aboriginals than those without complications and non-aboriginals respectively. The interactions with time show that the readmission hazard declines more rapidly with duration in the later years of the sample. This is consistent with the steeper slopes of the Kaplan Meier estimates for the last three years as shown in Figure 3.

I use the results from the second column of Table 5 to compute the estimated unconditional readmission rates at different durations for each of the four years. I report the unconditional readmission rate because it is more natural to think of readmission rates as unconditional than as hazard and it is possible to compare the estimated unconditional readmission rates with the actual readmission rates in Table 3a and 3b. The estimated readmission rates are obtained by computing the estimated survival rate for each patient at each duration, subtracting it from one and averaging over all patients. These estimated unconditional readmission rates are reported in Table 6a, 6b and 6c. Even though the readmission rates in Table 6 focuses only on seven, sixty and ninety days, the duration model allows the calculation of readmission rates for any number of days up to 90 days. The duration model allows all readmission rates to be computed from a single estimation. I compute the 60-day readmission rates because as already explained it takes that long for the body to return it the pre-pregnancy state.

Since there is no statistically significant difference between the readmission rates of Caesarean and vaginal deliveries, I classify the patients according to race and complications and compare with the actual readmission rates in Table 3a and 3b. The estimated readmission rates are similar to the actual readmission rates in Table 3a and 3b in size as well as in trend. The increased readmission rates over the years cannot readily be attributed to the policy because it could be due to some changes in the patient's characteristics. I use the estimates of the readmission coefficients to compute the effect of the policy. The similarities in the estimated and actual readmission rates imply a good estimation of the readmission regression and so these estimates can confidently be used to estimate the effect of the policy on readmission rates.

3.4.4 Effect of the Policy on Readmission Rates

To find the effect of the policy on readmission rates I first compute the unconditional readmission rates for each of the years after 1993 using the data from each year. For example for 1994, I use the 1994 data and the coefficients from the second column of Table 5, but without the year dummies or interaction variable coefficients, to compute the unconditional readmission rates. These readmission rates represent what the readmission rates would have been without the policy. I then subtract these from the readmission rates reported in Table 6. The difference between the two readmission rates represents the effect of the policy on the 1994 readmission rates, holding everything else constant. These are reported in Table 7a, 7b and 7b. Because I use data on the same patients to compute the two readmission rates, taking the difference eliminates the effect of the patients' characteristics on the readmission rates.

The numbers in Table 7a, b, and c shows the effect of the policy on the readmission rate in the relevant year. For example, the 1996 column shows the predicted readmission rate in 1996 had the 1993 policy remained in place. The first panel shows the weighted averages of the readmission rates when aboriginals and non-aboriginals are combined. In general the effect was smaller in 1995 than in 1994, and was bigger in1996. This pattern may imply that the high increase in readmission rates in 1994 probably prompted some short-term adjustment to improve the quality of care but the quality worsened again with time. The table shows that the effect of the policy increases with the days after discharge. The policy then had both short term and long-term detrimental effects on quality of care.

As already explained, if there is any effect of the reduction in length of stay on readmission rates it will be greatest on the readmission rates shortly after discharge. The effect of the policy on the 7-day readmission rate represents the short-term effect of the policy. These are shown in Table 7a. Even though there was no statistically significant difference in the reduction of length of stay for the categories of patients, the effect on readmission rates varies. The readmission rates for aboriginals increased more than those of non-aboriginals. In addition the effect is greater on the readmission rates of patients without complications than on those with complications. The greater effect of the policy on readmission rates of patients without complication is not surprising because given that both patients with and without complications go home early and both require home care, it is possible that the patients with complications got better home care supported by the Closer to Home Fund.

Table 7b and 7c show the cumulative effect of the policy on the 60-day and 90-day readmission rates is greater than those of the 7-day readmission rates. The greater cumulative effect on the 60-day and the 90-day readmission rates imply that the transfer of care from the hospitals had long term as well as short term effect on readmission rates. As already explained, the puerperium takes about sixty days and so the high cumulative effect on the 60-day readmission rates implies that the policy impeded the body's adjustment to the pre-pregnancy state. The higher cumulated effect of the policy on the 90-day readmission rates relative to the 60-day readmission rates could also imply that the policy affected the psychological adjustment of the mothers to childbirth. The long-term effect is also higher for aboriginals and patients without complications than non-aboriginals and patients with complications respectively.

3.5 Conclusion

This paper has shown that the transfer of care from hospitals to home succeeded in reducing hospital length of stay. Since the policy of early discharge reduces the hospital utilization rates, hospitals are able to reduce the cost of care. The Closer to Home Funds allows care to continue in the patient's home. In this way the patient receives the care she needs but the hospital does not have to pay for the housekeeping cost. This cost is borne by the patient. If it is cheaper for the patient's family to provide the housekeeping service than the hospital then the policy reduced the social cost of care, otherwise the policy simply transferred cost to the patient and did not improve social welfare.

The paper also shows that any cost saving that must have been achieved by the hospitals as a result of the policy must have come at the cost of the possible deterioration of the quality of care that patients received shown through the increase in readmission rates. The deterioration of the quality of care may be due to inadequate home care provided through public care. An improvement in the organization of the home care by the public care then could reduce such deterioration of health.

Statistically, the length of stay for delivery decreased by the same degree for both aboriginals and non-aboriginals. However, the policy must have deteriorated the quality of care of aboriginals more than non-aboriginals because the readmission rates of aboriginals increased more than those of non-aboriginals. The home care provision then may not have taken the tendency of aboriginals to have a poor health outcome into account. Further steps should be taken to improve the health outcome for aboriginals.

3.6 Reference List

British Columbia Vital Statistics (2001), "Selected Vital Statistics and Health Status Indicators", Annual Report, http://www.vs.gov.bc.ca/stats/annual/2001/tab01.html

British Columbia Reproductive Care Program (2002), *Report on the Findings of a Consensus Symposium on the Provision of Postpartum Services in British Columbia*, Retrieved June 20th, 2004 from http://www.rcp.gov.bc.ca/publications.htm

DesHarnais, S., E. Kobrinski, J. Chesney, M. Long, R. Ament, and S. Fleming (1987), "The Early Effects of the Prospective Payment System on Inpatient Utilization and the Quality of Care" *Inquiry*, 24(2), 7-16.

Gazmararian, J. A., J. P. Koplan (1996), "Length-of-Stay After Delivery: Managed Care Versus Fee-for –Service," *Health Affairs*, 15(4), 74-80.

Greene, W. H. (1993), *Econometric Analysis*, New York: Maxwell Macmillan International Publications Group.

Hansard (1995), "Proceedings in the Douglas Fir Room" *Official Report of the Legislative Assembly*, 20(1), http://www.legis.gov.bc.ca/1995/hansard/h0516am.htm.

Health Canada (2002), *Sharing the Learning. The Health Transition Fund, Synthesis Series, Aboriginal Health*, Ottawa: Publications Health Canada.

Kahn, K., W. Rogers, L. Rubenstein, M. Sherwood, E. Reinisch, E. Keeler, D. Draper, J. Kosecoff, and R. Brook. (1990), "Comparing Outcomes of Care Before and After Implementation of the DRG-Based Prospective Payment System," *Journal of the American Medical Association*, 264(15), 1984-73.

Keeler, E. B., K. L. Kahn, D. Draper, M. J. Sherwood, L. V. Rubenstein, E. J. Reinisch, J. Kosecoff and R. H. Brook (1990), "Changes in Sickness at Admission Following the Introduction of the Prospective Payment System" *Journal of the American Medical Association*, 264(15), 1962-68.

Kennedy, P. (1998), *A Guide to Econometrics*, Massachusetts: MIT Press.

Kosecoff, J., K. Kanh, W. Rogers, E. Reinisch, M. Sherwood, L. Rubenstein, D. Draper, C. Roth, C. Chew and R. Brook (1990), "Prospective Payment System and Impairment at Discharge: 'The Quicker and Sicker' Story Revisited," *Journal of the American Medical Association,* 264(15), 1980-83.

Manton, K. G., M. A. Woodbury, J. C. Vertrees and E. Stallard (1993), "Use of Medicare Services Before and After Introduction of the Prospective Payment System," *Health Services Research*, 28(3), 269-80.

McGrail, M.K., R. G. Evans, M. L Barer and S. B. Sheps, et al (2001), "The Quick and the Dead: Managing Inpatient Care in British Columbia Hospitals," *Health Services Research*, 35(6), 1319-38.

Rubenstein, L., K. Kahn, E. Reinisch, M. Sherwood, W. Rogers, Karnberg, D. Draper, and R. Brook (1990), "Changes in Quality of Care for Five Diseases Measured by Implicit Review, 1981-1986," *Journal of the American Medical Association*, 264(15), 1974-79.

Schrier, D and F. Ip (1994), *British Columbia's Changing Ethnic Mosaic*, Retrieved June 24, 2004 from http://www.bcstats.gov.bc.ca/data/pop/pop/ethnic.pdf

Tai-Seale, M., A.T. LoSasso, D.A. Freund and S.E. Gerber (2001), "The Long-Term Effects of Medicaid Managed Care on Obstetric Care in Three California Counties" *Health Services Research*, 36(4), 751-71.

Weinberger, M., D. M. Smith, B.P Katz, and P.S. Moore (1988), "The Cost-Effectiveness of Intensive Postdischarge Care," *Medical Care*, 26(11), 1092-1101.

Weissman, J.S., R.S. Stern, & A.M. Epstein (1994), "The Impact of Patient Socioeconomic Status and Other Social Factors on Readmission: A Prospective Study in Four Massachusetts Hospitals," *Inquiry*, 31, 163-72.

Welch, H.G., E.H. Larson, G. Hart and R.A Rosenblatt (1992), "Readmission after Surgery in Washington State Rural Hospitals," *American Journal of Public Health*, 82(3), pg 407-11.

3.7 Tables

Table 3.1 Sample Descriptions

(a) Sample Description

	Non-Aboriginals	Aboriginals	All
Percent of sample	96.8	3.2	100
Average age	29.5	24.9	29.4
Percent complication	65.4	63.3	65.3
Percent caesarean delivery	20.6	15.0	20.6
Percent readmissions	4.2	6.6	4.3
Percent of readmission sample	95	5	100

(b) Sample Description by Year, Deliveries

	1993	1994	1995	1996	All years
All					
Percent caesarean delivery	20.0	20.0	20.4	20.5	20.4
Percent complications	65.0	65.1	66.0	65.2	65.3
Total cases	23,148	23,325	23,538	22,583	92,594
Non-aboriginals					
Percent caesarean delivery	20.1	20.0	20.6	21.2	20.8
Percent complications	65.1	65.1	66.0	65.3	65.4
Total cases	22,376	22,607	22,795	21,895	89,673
Aboriginals					
Percent caesarean delivery	15.2	14.3	14.3	16.4	15.0
Percent complications	63.1	65.0	64.1	61.2	63.4
Total cases	772	718	743	688	2,921

Table 3.2: Average Length of Stay for Deliveries

	1993	1994	1995	1996	All years
All					
Caesarean delivery	5.9	5.6	5.5	5.3	5.6
Vaginal delivery	3.2	3.0	2.7	2.6	2.9
With complications	4.2	4.0	3.8	3.1	3.8
Without complications	3.6	3.3	3.1	3.1	3.3
Non-Aboriginals					
Caesarean delivery	5.9	5.6	5.5	5.3	5.6
Vaginal delivery	3.2	2.9	2.7	2.6	2.9
With complications	4.2	4.0	3.8	3.1	3.8
Without complications	3.6	3.3	3.1	3.1	3.3
Aboriginals					
Caesarean delivery	6.6	6.4	7.2	5.9	6.5
Vaginal delivery	3.4	3.2	3.0	3.0	3.2
With complications	3.4	3.6	3.6	3.9	3.6
Without complications	3.7	3.7	3.6	3.4	3.6

Table 3.3: Readmission Rates

(a) 7-day Readmission Rates (%)

	1993	1994	1995	1996	All years
All					
Caesarean delivery	1.1	1.5	1.5	1.2	1.3
Vaginal delivery	0.8	0.9	0.4	0.9	0.7
With complications	0.9	1.1	1.0	1.0	1.0
Without complications	0.8	0.9	0.8	1.0	0.9
Non-aboriginals					
Caesarean delivery	1.0	1.4	1.5	1.8	1.3
Vaginal delivery	0.7	0.8	0.4	0.9	0.7
With complications	0.8	1.0	1.0	1.0	1.0
Without complications	0.7	0.8	0.8	1.0	0.8
Aboriginals					
Caesarean delivery	3.1	4.0	0.9	1.8	3.0
Vaginal delivery	1.1	1.2	1.1	0.5	0.9
With complications	1.6	1.3	1.5	0.8	1.3
Without complications	1.4	2.0	1.1	0.8	1.3

(b) 90-day Readmission Rates (%)

	1993	1994	1995	1996	All years
All					
Caesarean delivery	4.6	4.8	4.9	5.5	5.0
Vaginal delivery	3.8	4.3	3.9	4.2	4.0
With complications	4.0	4.6	4.2	4.4	4.3
Without complications	4.1	4.1	3.9	4.5	4.1
Non-aboriginals					
Caesarean delivery	4.5	4.6	4.7	5.3	4.8
Vaginal delivery	3.8	4.3	3.8	4.1	4.0
With complications	4.0	4.6	4.1	4.3	4.2
Without complications	4.0	4.0	3.8	4.4	4.1
Aboriginals					
Caesarean delivery	11.1	12.7	14.1	12.4	12.5
Vaginal delivery	4.9	5.2	5.8	6.9	5.7
With complications	5.8	5.3	6.0	8.1	6.2
Without complications	5.9	7.9	7.6	7.1	7.1

Table 3.4: OLS Estimates, Delivery Length of Stay

	LOS	*LOS (with interactions)*
ABORIGINALS	0.320	0.304
	(0.00)	(0.000)
AGE	-0.134	-0.134
	(0.00)	(0.000)
*AGE*AGE*	0.002	0.002
	(0.00)	(0.000)
COMPLICATIONS	0.052	0.053
	(0.120)	(0.332)
CAESAREAN	2.695	2.645
	(0.00)	(0.000)
1994	-0.191	-0.166
	(0.00)	(0.000)
1995	-0.399	-0.405
	(0.00)	(0.018)
1996	-0.487	-0.544
	(0.00)	(0.000)
*COMPLICATIONS*1994*		-0.060
		(0.441)
*COMPLICATIONS*1995*		-0.016
		(0.837)
*COMPLICATIONS*1996*		0.076
		(0.308)
*ABORIGINALS*1994*		0.022
		(0.362)
*ABORIGINALS*1995*		-0.004
		(0.855)
*ABORIGINALS*1996*		0.002
		(0.564)
*CAESAREAN*1994*		0.106
		(0.184)
*CAESAREAN*1995*		0.095
		(0.220)
*CAESAREAN*1996*		0.069
		(0.662)
CONSTANT	4.581	4.971
	(0.00)	(0.000)
R^2	0.810	0.810
SAMPLE SIZE	92,594	92,594

P-values are in parenthesis

Table 3.5: Logit Estimates, Readmission Hazard

	γ_{iht}	γ_{iht} (*with interactions*)
ABORIGINALS	0.523	1.023
	(0.00)	(0.00)
AGE	-0.101	-0.096
	(0.00)	(0.00)
*AGE*AGE*	0.002	0.002
	(0.00)	(0.00)
COMPLICATIONS	0.102	2.553
	(0.023)	(0.00)
1994	1.131	3.442
	(0.00)	(0.00)
1995	1.022	3.467
	(0.00)	(0.00)
1996	1.107	3.704
	(0.00)	(0.00)
CAESAREAN	0.041	0.021
	(0.426)	(0.321)
T	-0.064	-0.045
	(0.00)	(0.00)
T2	0.210	0.120
	(0.150)	(0.340)
T^2	0.0001	0.0002
	(0.00)	(0.006)
*COMPLICATIONS*1994*		-2.506
		(0.00)
*COMPLICATIONS*1995*		-2.606
		(0.00)
*COMPLICATIONS*1996*		-2.654
		(0.00)
*ABORIGINALS*1994*		-0.691
		(0.008)
*ABORIGINALS*1995*		-0.556
		(0.033)
*ABORIGINALS*1996*		-0.492
		(0.054)
*T*1994*		-0.016
		(0.097)
*T*1995*		-0.017
		(0.075)

	γ_{iht}	γ_{iht} *(with interactions)*
*T*1996*		-0.032
		(0.001)
*T^2*1994*		0.0001
		(0.016)
*T^2*1995*		0.0001
		(0.021)
*T^2*1996*		0.0002
		(0.00)
CONSTANT	-6.07	-8.430
	(0.000)	(0.00)
SAMPLE SIZE	7,365,937	7,365,937

P-values are in parenthesis

Table 3.6: Estimated Unconditional Readmission Rates

(a) Estimated Unconditional 7-day Readmission Rates (%)

	1993	1994	1995	1996	All years
All					
With complications	0.6	1.4	1.0	1.2	1.1
Without complications	0.04	0.4	1.3	2.3	1.2
Non-aboriginals					
With complications	0.6	1.4	1.0	1.1	1.0
Without complications	0.04	0.3	1.3	2.1	0.9
Aboriginals					
With complications	1.5	1.7	1.9	2.2	1.8
Without complications	0.1	1.9	2.0	2.8	1.7

(b) Estimated Unconditional 60-day Readmission Rates (%)

	1993	1994	1995	1996	All years
All					
With complications	2.3	6.1	4.7	5.1	4.6
Without complications	1.2	5.2	4.2	4.6	3.9
Non-Aboriginals					
With complications	2.2	6.0	4.6	5.0	4.5
Without complications	1.2	5.1	4.1	4.5	3.7
Aboriginals					
With complications	5.4	9.8	7.1	9.6	7.9
Without complications	1.5	8.1	7.4	7.9	6.1

(c) Estimated Unconditional 90-day Readmission Rates (%)

	1993	1994	1995	1996	All years
All					
With complications	6.6	6.2	5.9	7.1	6.8
Without complications	3.0	5.8	5.5	8.2	5.6
Non-aboriginals					
With complications	6.6	6.0	5.7	7.0	6.3
Without complications	3.0	5.7	5.4	8.1	5.5
Aboriginals					
With complications	7.1	12.6	8.4	12.2	9.5
Without complications	4.1	9.9	7.5	10.9	6.1

Table 3.7: Effect of Policy on Readmission Rates

(a) Effect of the Policy on the Unconditional 7-day Readmission Rates (% points)

		1994	1995	1996	
All					
With complications		0.8	0.4	0.7	
Without complications		1.3	1.3	2.2	
Non-aboriginals					
With complications		0.8	0.4	0.5	
Without complications		1.3	1.2	2.1	
Aboriginals					
With complications		0.2	0.4	3.7	
Without complications		1.8	1.9	5.7	

(b) Effect of Policy on the 60-day Readmission Rates (% points)

		1994	1995	1996	
All					
With complications		3.9	1.9	2.8	
Without complications		4.9	2.6	4.4	
Non-aboriginals					
With complications		3.8	1.9	2.8	
Without complications		4.9	2.4	4.3	
Aboriginals					
With complications		4.3	3.2	4.0	
Without complications		7.7	7.0	7.5	

(c) Effect of Policy on the 90-day Readmission Rates (% points)

		1994	1995	1996	
All					
With complications		3.1	2.6	6.5	
Without complications		5.6	5.7	8.1	
Non-aboriginals					
With complications		3.1	2.5	6.4	
Without complications		5.5	5.5	7.9	
Aboriginals					
With complications		1.2	5.0	8.4	
Without complications		7.9	12.0	10.3	

3.8 Figures

Figure 3.1: Length of Stay

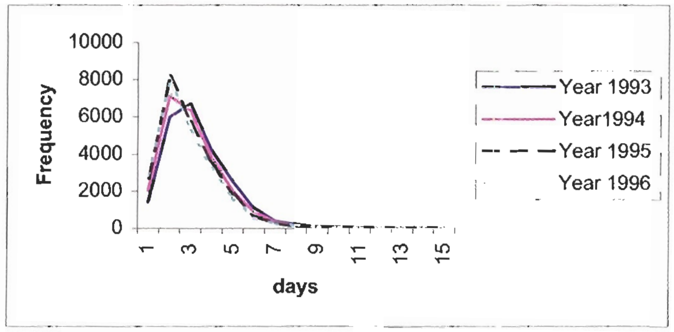
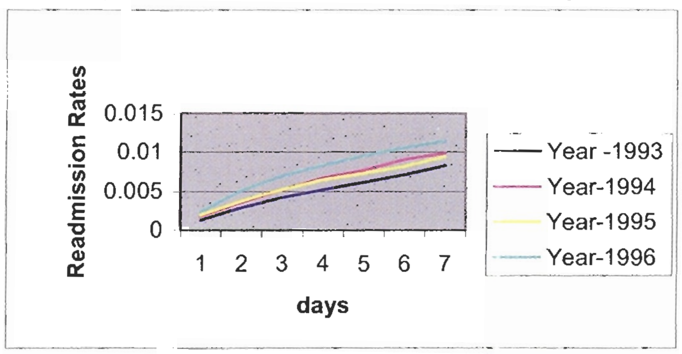


Figure 3.2: Readmission Rates

(a) Cumulative Readmission Rates for 7 days



(b) Cumulative Readmission Rates

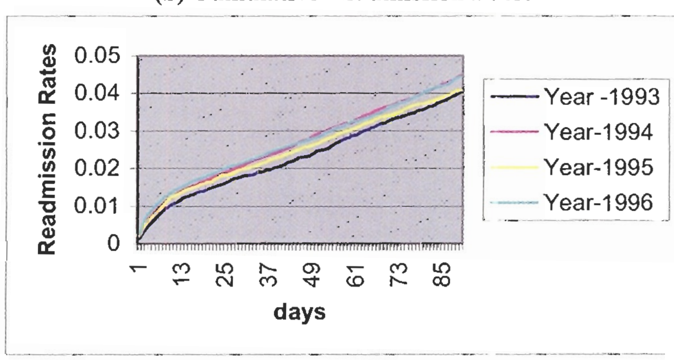
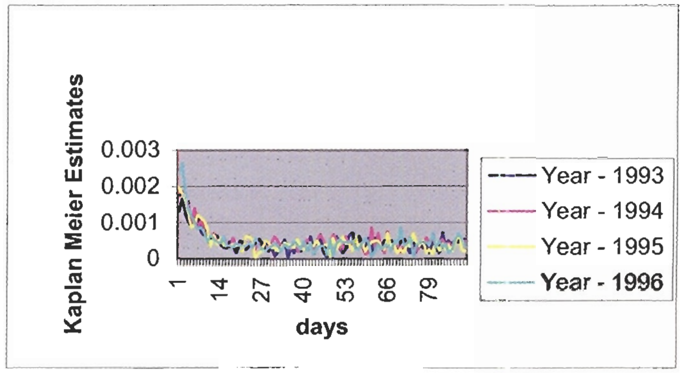


Figure 3.3: Kaplan Meier Estimates for the Readmission Hazard



BIBLIOGRAPHY

Andreoni, J. (1991), "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?" *Rand Journal of Economics* 22, 385-95.

Becker, G.S. (1968), "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, 169-217.

British Columbia Vital Statistics (2001), "Selected Vital Statistics and Health Status Indicators", Annual Report, http://www.vs.gov.bc.ca/stats/annual/2001/tab01.html

Blomqvist, A. and H. Horn (1984), "Public Health Insurance and Optimal Income Taxation", *Journal of Public Economics*, 24, 353-71.

British Columbia Reproductive Care Program (2002), *Report on the Findings of a Consensus Symposium on the Provision of Postpartum Services in British Columbia*, Retrieved June 20th, 2004 from http://www.rcp.gov.bc.ca/publications.htm

Burbridge, J.B. and G.M. Myers (2004), "Tariff Wars and Trade Deals with Costly Government," *Discussion Paper*, Department of Economics, Simon Fraser University.

Butler, R.J., D.L. Durbin, and N.M. Helvacian (1996), "Increasing Claims for Soft Tissue Injuries in Workers' Compensation: Cost Shifting and Moral Hazard," *Journal of Risk and Uncertainty*, 13(1), 73-87.

Butler, R.J, R.P. Hartwig and H. Gardner (1997), "HMOs, Moral Hazard and Cost Shifting in Workers' Compensation," *Journal of Health Economics*, 16, 191-206.

Chalkley M. & J.M. Malcomson (1995), "Contracting for Health Services with Unmonitored Quality," *Discussion Paper*, Department of Economics, University of Southampton.

DesHarnais, S., E. Kobrinski, J. Chesney, M. Long, R. Ament, and S. Fleming (1987), "The Early Effects of the Prospective Payment System on Inpatient Utilization and the Quality of Care" *Inquiry*, 24(2), 7-16.

Dor, A. and D.E. Farley (1996), "Payment Source and the Cost of Hospital Care: Evidence from Multiproduct Cost Function with Multiple Payers," *Journal of Health Economics*, 15(1), 1-21.

Dranove, D. (1988), "Pricing by Non-Profit Institutions: Case of Hospital Cost Shifting," *Journal of Health Economics*, 7(1), 47-57.

Ellis, R.P. & T.G. McGuire (1986), "Provider Behaviour Under Prospective Reimbursement, Cost Sharing and Supply," *Journal of Health Economics*, 5, 129-51.

Ettner, S. (1996), "New Evidence on the Relationship between Income and Health," *Journal of Health Economics,* 15, 67-85.

Evans, R.G. (1984), *Strained Mercy: The Economics of Canadian Health Care,* Toronto: Butterworths.

Fuchs, V.R. (1975), *Who Shall Live? Health Economics and Social Choice*, Expanded Edition, New York: Basic Books.

Gazmararian, J. A., J. P. Koplan (1996), "Length-of-Stay After Delivery: Managed Care Versus Fee-for –Service," *Health Affairs*, 15(4), 74-80.

Gertler, P.J. (1989), "Subsidies, Quality, and the Regulation of Nursing Homes," *Journal of Public Economics,* 38, 33-52.

Greene, W. H. (1993), *Econometric Analysis*, New York: Maxwell Macmillan International Publications Group.

Hanchak, N., N. Schlackman, & S. Harmon-Weiss (1996), "U.S. Healthcare's Quality-based Compensation Model," *Health Care Financing Review,* 17, 143-59.

Hansard (1995), "Proceedings in the Douglas Fir Room" *Official Report of the Legislative Assembly,* 20(1), http://www.legis.gov.bc.ca/1995/hansard/h0516am.htm.

Health Canada (2002), *Sharing the Learning. The Health Transition Fund, Synthesis Series, Aboriginal Health*, Ottawa: Publications Health Canada.

Kahn, K., W. Rogers, L. Rubenstein, M. Sherwood, E. Reinisch, E. Keeler, D. Draper, J. Kosecoff, and R. Brook. (1990), "Comparing Outcomes of Care Before and After Implementation of the DRG-Based Prospective Payment System," *Journal of the American Medical Association*, 264(15), 1984-73.

Keating, B (1984), "Cost Shifting: An Empirical Examination of Hospital Bureaucracy," *Applied Economics*, 16(2), 279-89.

Keeler, E. B. (1990), "What Proportion of Hospital Cost Differences is Justifiable?," *Journal of Health Economics*, 9(3), 359-65.

Keeler, E. B., K. L. Kahn, D. Draper, M. J. Sherwood, L. V. Rubenstein, E. J. Reinisch, J. Kosecoff and R. H. Brook (1990), "Changes in Sickness at Admission Following the Introduction of the Prospective Payment System" *Journal of the American Medical Association*, 264(15), 1962-68.

Kennedy, P. (1998), *A Guide to Econometrics*, Massachusetts: MIT Press.

Kesteloot, K. & N. Voet (1998), "Incentives For Cooperation in Quality Improvement Among Hospitals – the Impact of the Reimbursement System," *Journal of Health Economics*, 17, 701-28.

Kosecoff, J., K. Kanh, W. Rogers, E. Reinisch, M. Sherwood, L. Rubenstein, D. Draper, C. Roth, C. Chew and R. Brook (1990), "Prospective Payment System and Impairment at Discharge: 'The Quicker and Sicker' Story Revisited," *Journal of the American Medical Association,* 264(15), 1980-83.

Laffont, J. J. and Tirole, J. (1989), "Using Cost Observation to Regulate Firms," *Journal of Political Economy,* 94(3, Part 1), 614-41.

Lee, J.A., H. Birnbaum, H., & C. Bishop (1983), "How Nursing Homes Behave: A Multi-Equation Model Nursing Home Behaviour," *Social Science Med.,* 17, 1897-1906.

Leger, P. T. (2000), "Quality Control Mechanisms Under Capitation Payment for Medical Services," *Canadian Journal of Economics,* 33(2), 565-86.

Little, J.S (1992), "Public-Private Cost shifts in Nursing Home Care," *New England Economic Review*, 0(7), 3-14.

Ma, C.A (1998), "Cost and Quality Incentives in Health Care: Altruistic Providers," *Discussion Paper,* Department of Economics, Boston University.

Ma, C.A. & T.G. McGuire (1997), "Optimal Health Insurance and Provider Payment," *American Economic Review*, 87(4), 685-704.

Manton, K. G., M. A. Woodbury, J. C. Vertrees and E. Stallard (1993), "Use of Medicare Services Before and After Introduction of the Prospective Payment System," *Health Services Research*, 28(3), 269-80.

McDonough, G. J., D. Duncan, Williams and J. House (1997), "Income Dynamics and Adult Mortality in the United States, 1972 through 1989," *American Journal of Public Health,* 87,(9) 1476-83.

McGrail, M.K., R. G. Evans, M. L Barer and S. B. Sheps, et al (2001), "The Quick and the Dead: Managing Inpatient Care in British Columbia Hospitals," *Health Services Research,* 35(6), 1319-38.

Morrisey, M.A. and F.A. Sloan (1989), "Hospital Cost Shifting and the Medicare Prospective Payment System," *Working Paper*, University of Alabama at Birmingham.

Morrisey, M.A. (1995), "Movies and Myths: Hospital Cost Shifting," *Business Economics*, 30(2), 22-25.

Norton, E. (1992), "Incentive Regulation of nursing Homes," *Journal of Health Economics*, 11, 105-28

Nyman, J. (1985), "Prospective and 'Cost-plus' Medicaid Reimbursement, excess Medicaid Demand, and the Quality of Nursing Home Care," *Journal of Health Economics,* 4, 237-59.

Pauly, M. (1980), *Doctors and their Workshops*, Chicago: U. of Chicago Press.

Petretto, A. (1999), "Optimal Social Health Insurance with Supplementary Private Insurance," *Journal of Health Economics*, 18, 727-45.

Pope, G. (1990), "Hospital Nonprice Competition and Medicare Reimbursement Policy," *Journal of Health Economics*, 8, 147-72.

Rickman, N. and A. McGuire (1999), "Regulating Provider's Reimbursement in a Mixed Market for Health Care," *Scottish Journal of Political Economy*, 46(1), 53-71.

Rubenstein, L., K. Kahn, E. Reinisch, M. Sherwood, W. Rogers, Karnberg, D. Draper, and R. Brook (1990), "Changes in Quality of Care for Five Diseases Measured by Implicit Review, 1981-1986," *Journal of the American Medical Association*, 264(15), 1974-79.

Shavell, S. (1991), "Specific versus General Enforcement of Law," *Journal of Political Economics*, 99, 1088-1109.

Schleifer, A.(1985), "A Theory of Yardstick Competition," *The RAND Journal of Economics,* 16(3), 319-27.

Schrier, D and F. Ip (1994), *British Columbia's Changing Ethnic Mosaic*, Retrieved June 24, 2004 from http://www.bcstats.gov.bc.ca/data/pop/pop/ethnic.pdf

Showalter, M. H. (1985), " Physician's Cost Shifting Behaviour: Medicaid versus other Patinets," *Contemporary Economic Policy*, 15(2), 74-84.

Sloan, F. A. and E. R Becker (1984), "Hospital Ownership and Performance," *Economic Inquiry,* 23(1), 21-36.

Tai-Seale, M., A.T. LoSasso, D.A. Freund and S.E. Gerber (2001), "The Long-Term Effects of Medicaid Managed Care on Obstetric Care in Three California Counties" *Health Services Research*, 36(4), 751-71.

Weinberger, M., D. M. Smith, B.P Katz, and P.S. Moore (1988), "The Cost-Effectiveness of Intensive Postdischarge Care," *Medical Care*, 26(11), 1092-1101.

Weissman, J.S., R.S. Stern, & A.M. Epstein (1994), "The Impact of Patient Socioeconomic Status and Other Social Factors on Readmission: A Prospective Study in Four Massachusetts Hospitals," *Inquiry*, 31, 163-72.

Welch, H.G., E.H. Larson, G. Hart and R.A Rosenblatt (1992), "Readmission after Surgery in Washington State Rural Hospitals," *American Journal of Public Health*, 82(3), pg 407-11.

Wyszewianski, L., J. Thomas, & B. Friedman (1987), "Case-based Payment and the Control of Quality and Efficiency in Hospitals," *Inquiry*, 24, 17-25.

Zuckerman, S. (1987), "Commercial Insurers and All-Payer Regulation, Evidence on Hospital's Responses to Financial Need," *Journal of Health Economics*, 6, 165-87.