UNIVERSITY OF SOUTHAMPTON

FACULTY OF LAW, ARTS & SOCIAL SCIENCES

School of Social Sciences

Health Investment: How can the state influence the consumer to obtain the maximum health investment?

by

Manpreet Kaur Sidhu

Thesis for the degree of Doctor of Philosophy

October 2006

UNIVESITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF LAW, ARTS, AND SOCIAL SCIENCES
SCHOOL OF SOCIAL SCIENCES

DOCTOR OF PHILOSOPHY

HEALTH INVESTMENT: HOWCAN THE STATE INFLUENCE THE CONSUMER TO OBTAIN THE MAXIMUM HEALTH INVESTMENT?

by MANPREET KAUR SIDHU

This thesis examines the factors that influence the individual's decision to undertake an investment in health that, in the short term, provides a health improvement and, in the longer term, provides a potential screening benefit which may require additional health investment. Additionally the thesis examines the methodology of modelling tools that support decision-makers in making resource allocation decisions.

The main hypothesis examined is 'what factors affect the individual's decision to undertake an initial health investment and what factors affect the individual's decision to undertake the secondary investment, which incurs a further cost', this is examined as a number of separate questions.

The first chapter reports the literature review of economic theory surrounding health investment especially in relation to prevention, which is applied to the sight test; examining the sight test as a health improvement and as a screening tool. The second chapter empirically examines 'what factors impact the individual's decision to undertake the initial investment and what factors impact the individual's decision to undertake a re-investment in health, based on an increase in information', specifically in the sight test market. The third chapter empirically examines the second part of the main hypothesis, 'how does the risk of requiring an additional health investment impact on the decision to undertake the initial investment', specifically within the general practitioner and prescription charges paradigm. The fourth chapter focuses upon the additional analysis examining the methodology of modelling tools used within healthcare resource allocation decision-making where resources are limited and the impact upon final decision of using a specific tool.

It is found that the risk of requiring an additional (costly) investment will decrease the likelihood of initial investment in health, though once the initial investment is undertaken additional investment will be undertaken if required. The choice of modelling methodology will, given identical data and assumptions, not affect the final outcome or alter the decision-makers final result.

There is a gap in the literature when assessing what factors affect the uptake of an initial health investment (zero cost) when a follow-up investment may mean a cost to the individual, which this thesis adds to. There is also a gap in literature on an overall examination (in one place) of the impact of different resource allocation tools upon actual resource allocation decisions, the thesis provides the non-expert decision-maker with a guide to assess the most appropriate tool to use in a given situation.

TABLE OF CONTENTS

Acknowledgements

I would like to acknowledge the support and guidance of my supervisor Dr Patricia Rice in providing constant supervisory meetings in order to make sure I got on with it, and her comments and re-comments on my work that have helped me produce my final completed thesis.

I would like to thank my employers, current and previous, in providing support and understanding in allowing me to reach my goal.

I would like to thank my family and friends for the constant support and encouragement. I would like to thank Dr Angela Reidy and Dr Parul Desai, without them I never would have started on this 'learning path'.

Most of all I would like to thank my Mum and sister, Dawandeep, in providing constant support, never letting me give up and helping me achieve my dream; without them both I would have given up a long time ago.

1 Introduction

What factors influence the individual's decision to undertake an investment in health that, in the shorter term, provides a health improvement, and, in the longer term, provides a potential screening benefit which may require additional health investment.

Health is an important part of the individual's health production function, if the individual is in poor health s/he will consume more of society's resources, health care, and welfare. An individual in poor health will not operate at his/her maximum production frontier, therefore will not be maximising his/her utility as well causing externalities to other consumers – individuals in poorer health consume more welfare resources and are more likely to affect consumption decisions made by other individuals. Consumers are most likely unaware that they are not operating at maximum health or that investment today will not only improve their health today but also reduce costs to themselves and the state tomorrow.

The types of health intervention examined in this thesis are two-pronged; they provide current health improvement as well as future health investment. A screening tool that screens for disease that is not life threatening whilst also providing a direct health improvement has not been assessed within economic literature. The most obviously connected area of work where this could be assumed to have been conducted is within the dental arena. Literature related to the economics of dental care (Manning and Phelps 1979)

relates more to the role of insurance in dental attendance, full and partial coverage of dental care tends towards an increase in dental attendance.

There are essentially two types of screening, opportunistic screening and screening programmes. Screening programmes such as cancer screening provide an obvious direct benefit to individuals, information regarding a disease that causes loss of life and huge disutility; however opportunistic screening for other 'non-deadly' diseases that do not cause immediate disutility or loss of life need to be examined in order to assess whether the cost is worth the benefit gained. The state needs to assess the most efficient way to encourage the individuals to attend for a health intervention that will improve future benefits and reduce future costs.

The hypothesis to be examined is that a screening tool that is free at point of contact to a group of individuals will not necessarily lead to increased uptake given that there is a potential future cost to the patient. In order to investigate our hypothesis, we examine the issue as a number of separate questions.

The initial question is: what factors affect the individual's decision to undertake the initial health investment and what factors affect the individual's decision to undertake the secondary health investment, which incurs a further cost.

The thesis begins with the theoretical section (paper one); a literature review of the economic theory surrounding health investment especially in relation

to prevention, this is then applied to the sight test; examining the sight test as a health improvement and as a screening tool.

Two sets of literature are currently available within this field, Grossman – relating to the health improvement aspect of a health good (the individual makes an investment to improve/increase health capital) and Phelps – relating to the preventative care aspect of a health tool (individual attends for a specific preventative intervention rather than acute medical care). Grossman finds that a price set to zero will not necessarily impact upon demand for health improvement but other factors (such as income and education) will impact upon the demand. Phelps finds that a reduction in the price of preventative care essentially reduces the demand for acute medical care. Though both papers examine health improvement and investment they are doing so disjointedly. Overall there is literature available assessing the theory of health investment, however there is a gap within the literature of the factors important in undertaking a health investment, specifically a health investment where a follow-up (potentially) costly investment maybe required; especially where the initial investment may have a zero cost to the individual but the follow-up investment requires a cost.

Following the theoretical section the thesis examines the hypothesis empirically in paper two; what factors impact the individual's decision to undertake the initial investment and what factors impact the individual's decision to undertake a re-investment in health, based on an increase in information; this is where the thesis tries to bridge the gap of knowledge and

not only examine the impact of socio-economic factors upon the demand for a health improvement tool but also the impact of those factors upon the preventative care that is produced by the same good. The impact of factors upon the uptake of sight testing and investment in vision has previously not been examined, especially relating to the impact that a zero price will have upon sight test uptake, though there has been analysis of the impact of taking eligibility away, ie when price is greater than zero. Paper two examines the impact of factors (such as eligibility) upon sight test attendance, and finds that initial attendance for a sight test is significantly less likely (-0.53) if the individual is not eligible for a free sight test. However, eligibility has no impact upon whether the individual will return for a repeat sight test two years after the initial attendance.

In order to examine the second part of the equation: how does the risk of requiring an additional health investment impact on the decision to undertake the initial investment? This is undertaken by examining the impact which prescription charges (as they increase year on year) and eligibility for free prescriptions have upon GP attendance and the impact of charges upon dispensation, not utilisation, of prescriptions in paper three.

This analysis examines the factors that potentially affect the individual's decision to undertake a health investment when the initial free investment is followed by a secondary investment cost. The health investment process is a three stage process, the individual will make the first investment (attend the GP), thereafter the GP makes a decision to provide the individual with a

prescription or not (in this element the individual has no real influence); the third element is the individual's decision to dispense the prescription within this stage the individual needs to make a further investment in medication which may or may not require a direct cost (ie prescription charge), the individual's decision to make this investment is based upon the information that s/he has gained from the GP.

The third paper firstly looks at the literature surrounding the effect of prescription charges upon prescription demand and utilisation, followed by an analysis of the effect of prescription charging upon GP attendance, and prescription giving, what factors affect behaviour at each level of the investment process. Literature are available that examine the impact of prescription charges upon prescription utilisation, which do not purport to answer the question we have in mind. There is minimal literature that examines the impact of prescription charges upon GP attendance; however the impact of charges upon GP attendance is very inconclusive. This thesis not only tries to answer the question how does the price of a potential additional health investment impact upon the initial decision to undertake the investment within the GP-prescription framework but also examines whether the perceived effect of prescription charges is not affected by another factor (ie the GP). The impact of factors upon the uptake of an initial investment that could lead to a further health investment, examined as the impact of prescription charges upon GP consultations and prescription dispensation, has not been examined within the literature. In this thesis paper three examines the impact of factors (such as eligibility and charges for

prescriptions) upon GP attendance, GP prescription writing and patient prescription dispensation, and finds individuals significantly less likely (-0.065) to attend for a GP consultation if they are liable to pay for a prescription charge and significantly more likely (0.27) to attend for a GP consultation if they are eligible for a free prescription due to social (ie income, employment-based) reasons.; whereas GP prescription writing and patient prescription dispensation are not affected by these factors.

Given that individuals' health investment decisions are dependent upon the level of insurance provided by the state, which in turn is dependent upon the limited resources available, it is necessary to assess the best way to maximise the health benefits given the limited resources available. There are a number of mathematical methods available, especially within the economic modelling arena; therefore an analysis of the most appropriate method of assessing and making these health investment decisions from the perspective of the state and budget holder needs to be undertaken. Therefore the final question is: what is the most appropriate tool available to the decision-maker, local or national, in order to direct limited resources (subject to a budget constraint) to maximise benefits?

The fourth paper examines the methodology surrounding resource allocation within the health framework, where resources are limited. Given the tools available is there a methodology to decide the best tool to use within a certain decision. This final paper in the thesis is not a consumer-driven paper but examines the provider perspective when making resource allocation decisions

within the NHS, NICE, the SMC and the AWMSG. Decision-makers are required to make decisions upon the use of health technologies to treat a certain disease however the decision is based upon a number of factors such as not only clinical effectiveness but also cost implications. Therefore modelling cost-effectiveness of these technologies is an important part of the decision process. Different modelling techniques are available and therefore it is important to understand the differences between each of these methodologies as well as the sensitivity of the outcomes (produced by these models) to the data applied. The impact that the three resource allocation tools have upon the actual resource allocation decision has not been assessed, though an assessment of any two of the three modelling techniques has been carried out. However an assessment of all three techniques (especially in terms of accuracy) in application to only one problem has not been undertaken. This paper provides a guide to the non-expert decision-maker, in order to assess the most appropriate tool to use in a given situation; is the marginal investment for the more complex model worth the marginal benefit, in terms of accuracy, justifiable?

The overall conclusion that can be extracted from this thesis is that to increase the individual's health investment and reduce potential future cost to society it is important that the marginal cost of the investment is equal to the private marginal benefit; the private marginal benefit must equal the social marginal benefit. A reduction in price of the investment can lead to an increase in demand however the shadow price of health is an important factor for demand. A two-pronged health investment is affected by price. If price

is equal to zero for the initial investment then the findings of our research show that investment will occur. Initial investment is contingent on the perception of requiring an additional investment, if the additional investment has a price equal to zero this will increase the likelihood of initial investment. The risk of requiring an additional costly investment will reduce the likelihood of the individual undertaking the initial investment; however once the initial investment is undertaken the cost of the additional investment is not a deterrent to the individual undertaking that additional investment.

The risk of the an additional secondary investment cost will decrease the likelihood of an initial investment in health, though once the individual has made the initial investment they will follow through with further investment if required; so it maybe worth carrying out a cost-benefit analysis of the impact of risk of requiring glasses and the uptake of the over 60s.

The choice of modelling methodology, given identical data and assumptions, will not affect the final outcome or alter the decision-makers final result.

Therefore the contribution made by this thesis is to add to the current knowledge on the factors and processes used by individuals to invest in their health given that the initial 'zero cost' health investment they are making may lead them to further cost. The thesis adds to knowledge by reporting that zero price is a significant incentive to getting individuals to make the initial health investment via screening, however a zero price does not impact on the individual making a repeat health investment via screening. The risk

of having to make a secondary non-zero price health investment will deter individual from making the initial zero-price health investment, however this thesis discovers that once the initial (zero price) health investment has been made any additional cost that the patient faces arising from the initial (zero price) health investment will be undertaken. This is a significant contribution to knowledge suggesting, specifically in the case of GP-prescriptions, and sight test – spectacles, that the secondary cost may deter initial investment but the patient follow through is there. The second part of the thesis, examining the three decision-making tools add significantly to knowledge by going beyond the analysis conducted by Karnon (2003), where the comparison was between markov and discrete event simulation, and finds that a comparison between markov and decision tree, and decision tree and discrete event simulation, using the same data and assumptions produces the same end result regardless of methodology used; thus extending the conclusion that Karnon (2003) arrived at with markov and discrete event comparison.

The analyses conducted and reported in this thesis essentially informs policy-making by recognising that individual's may not be maximising the full potential of a free health intervention (specifically the sight test) given that they are potentially faced with a very costly secondary investment (spectacles). If government wants to encourage individuals to undertake health investments it is necessary to examine all the factors affecting this decision, specifically in the case of the sight test it is relevant to offer free sight testing, as a method of opportunistic screening, to individuals over 60.

When making policy decisions it is important for the government to recognise that in order to encourage individuals to fully maximise the benefits of an opportunistic screening program, and the state to maximise uptake, the detrimental impact of the 'second' cost must be considered. It is not enough to simply provide a 'zero price' on the 'relatively' cheaper initial investment, whilst individuals face an expensive secondary cost, and expect uptake to be maximised.

2 Economic theory of a screening tool offering additional health improvement benefits: the sight test as an example

2.1 Introduction

This paper will try to examine the general issues that arise with provision of a tool that not only provides a general health improvement but also provides screening for the detection of a more serious issue. A specific example of such a tool is the sight test, which is examined in this paper.

The NHS provides a number of health services that offer a longer term health benefit such as screening for disease, however there are also health services available that provide not only a longer term screening benefit but also a shorter term health improvement.

Primary screening programmes, such as for instance breast or cervical cancer screening offer the patient and society a longer term benefit that is early disease detection leading to expected reduction in future costs and expected increase in future benefits (reduced morbidity and mortality) and are currently provided free at point of contact by the NHS. These types of screening programmes however do not offer the individual a direct health improvement, attending for a breast cancer screen does not improve anything straightaway (aside from providing a certain level of assurance), there is no tangible health gain.

However opportunistic screening, such as the sight test, is a tool that not only provides a benefit in terms of disease detection but also offers an additional benefit – a private (health improvement) benefit. The sight test is not a screening programme but does offer the opportunity to detect serious sight threatening disease whilst also offering the individual the incentive of a health improvement. As stated before primary screening programmes are provided free on the NHS, it is necessary to assess whether opportunistic screening programmes (ie the sight test) where a health improvement and screening benefit are both provided, should also have free provision.

There are two economic arguments for the provision and pricing of a 'two-pronged health service', efficiency and equity, which can be set out using the marginal cost (MC) and marginal benefit argument. The marginal private benefit (MPB) represents the individual's demand curve. The efficiency argument states that the individual will demand the health service if the MPB is greater than or equal to MC. The MPB is based upon the short term health improvement that the individual will gain, in the case of the sight test this is the improvement in vision (via a lenses prescription) that could occur after an optician visit. Government intervention due to efficiency can occur when the MPB is less than the marginal social benefit (MSB), this can occur because the individual underestimates his/her MPB, and also because externalities that occur due to health service consumption. In the case of the sight test individuals could be consistently underestimating their MPB due to a lack of information regarding the screening benefits of the sight test for illnesses such as glaucoma; providing further information regarding the screening

benefits could help to rectify this problem to some extent. Government intervention can also occur due to equity reasons the sight test can also provide additional social benefits via merit good qualities, ie improved vision or earlier treatment for sight illnesses can mean a decreased likelihood of accidents and reduced burden upon others, potentially reducing the impact on external resources. Equity argument for the sight test applies very much to the detection of glaucoma, a disease that affects vision and has a higher incidence in certain demographic groups (such as individuals of black ethnic origin over 40 years old).

In this context the benefits provided by an opportunistic screening tool (providing a health improvement as well) are the private benefit (provided by the health improvement), the expected private benefit is provided by the screening and the expected social benefit is provided by the screening. Any pricing rules need to consider all these three benefits in order to provide an efficient service.

An important question is why does the health service, and in specific this type of health service, require government intervention. For free health service provision, health services should be defined as a public good; but health care is not a public good. A public good is defined as '.....one which all enjoy in common in the sense that each individual's consumption of such a good leads to no subtraction from any other individual's consumption of that good' Cullis and Jones (1992) (p.60). Public goods are different to private goods, a private good can be given to an individual and

that sum can be subtracted from the total good available to other individuals in the economy.

The perfectly competitive market is the allocation and production mechanism that is deemed the 'gold standard' in economic theory where no intervention by the government occurs. However intervention in the market will generally occur when the standard assumptions (perfect information, perfect competition and, no market failures) do not hold and can only be corrected or improved by intervention.

Perfect information in the free market is important for both producers and consumers, in order for the market to work accurately. There is a need to intervene in the market and provide information if the knowledge that an activity is harmful may reduce consumption, an example cited by Cohen and Henderson (1991) is smoking. Smoking is a hazard to health, however even after an immense amount of publicity regarding these health risks, there has been little change in attitude[1]. The unregulated market for cigarettes will not provide the information needed to make informed choices, and consumers can not learn from experience, which provides a case for public provision of information.

If the individual does not have perfect information in order to the make an optimum decision and maximise his/her utility, then the state will intervene in order to provide the incentive for the individual to undertake the utility

[1] Cohen and Henderson (1991b) p.29

maximising decision. The methods of state intervention can vary between simply increasing the level of information available to the individual (ie advertising) or by direct intervention (ie taxation, subsidisation). Increasing the level of information available to the individual will allow him/her to maximise his utility, direct intervention by the state does not increase the individual's information but makes him/her more likely to undertake a utility maximising decision. With respect to the sight test the individual does not have perfect information of his/her own vision; the individual will have knowledge regarding the level of their vision and whether their sight has deteriorated; however s/he will not have knowledge regarding the potential diseases that can be detected.

The extent of perfect competition in the sight test market can be deemed to be regulated by the FODO, who establish the quality of the optometrists themselves as well as the prices and quality of the glasses and tests performed. However, since the deregulation of the market in 1994, there has been a significant step towards separating prescription-givers from prescription-providers in terms of glasses, in order to reduce the potential for supplier-induced demand. Individuals can also purchase reading glasses over the counter. Barriers to entry do exist within this market especially in terms of qualifications needed to become an optometrist.

The perfect market assumes that there are no externalities. In the free and perfect market, there is the assumption that all utility (both positive and negative) is yielded to the consumer alone, and all costs are borne by the

consumer alone, however this is not always the case. Individuals may be making decisions that are sub-optimal since they are only making decisions based upon their own costs and utilities. The state maybe left worse off after individuals make their decisions so making it necessary for the state to intervene. External resource implications of individual decisions in health include health care costs in terms of the NHS, output loss through sickness leave and external intangible costs. For consumption to be fully informed it is important that a rational decision is made based upon a comparison of marginal utility and cost, as per economic theory.

The sight test, or rather the non-attendance at the sight test, by individuals over 60 has externalities. Individuals not attending for a sight test especially when they are at an increased risk of suffering from a problem, such as glaucoma, where damage is irreversible, are more prone to accidents and more dependent upon others in order to get around thus giving rise to greater external resource implications.

As mentioned before there are equity reasons for state intervention within the sight test market, the sight test has possible merit good qualities that is it's consumption can lead to increased utility for other individuals. Increased consumption of the sight test by the over 60s will, in the long run, be beneficial to friends and family since they will not necessarily be required to assist in day to day activities, as may be the case if individuals are suffering from irreversible sight loss or blindness from glaucoma. Screening also leads to a social benefit by reducing the future cost to the NHS.

The sight test is different from other screening interventions since it provides a short term health improvement to the individual. By visiting an optometrist the individual will gain information regarding the state of their eyesight. It is possible from a sight test to check whether the individual is suffering from a reduction in vision, if they need glasses or lenses. The information that an individual receives from a visit to the optician is not only how much of the individual's sight has changed, but s/he will also extract information regarding colour vision deficiency, eye movement and co-ordination, whether a change (introduction) to glasses/lenses will improve the quality and clarity of vision[2], also the individual will be informed whether s/he has any other vision problems, such as myopia, astigmatism[3] which are among a number of problems (visual) that the individual could suffer from. Better sight can lead to an improvement in quality of life, and can also be safer.

Above are benefits that accrue to the individual from visiting the optometrist. However, there are external effects, or externalities, accruing if the individual does (not) visit the optometrist. If the individual does not ever visit the optician, then s/he will not know if their eyesight is weakening therefore they will not be able to improve their health in the short term. From a societal perspective sight threatening and potentially costly diseases will not be detected causing future costs to be higher and benefits lower. Screening can lead to future saving to the NHS.

[2] Colleg (1999)
[3] Colleg (1998b)

Thus, given that the sight test is not a pure public good, public sector intervention can be justified if there is some increase in the welfare of society that would not have been possible if left to the private sector. If the sight test is left to the private sector then only individuals whose marginal cost is in equilibrium with their marginal benefit will attend for a sight test, inevitably this will mean that from a societal perspective the marginal cost will not equal the marginal social benefit since individuals whose attendance is required for screening will not attend as their private marginal benefit is not equal to the social marginal benefit, individuals may value future benefits much lower than present benefits.

As stated previously the sight test is an example of the health service that provides a screening benefit and provides a health improvement. There are two aspects of the sight test; a tool to test and improve vision as well as a mechanism by which to screen individuals for a number of eye diseases.

The year 1989 saw a change in the way vision testing was paid for in the UK, with the introduction of sight test fees for everyone (unless they were eligible to be exempt) where there had previously been none. The eligibility of an individual, for a free NHS sight test, is dependent upon various health and income factors. In the year 1999, there was another change in the eligibility criteria allowing free sight tests for the over 60s.

The government may well feel the need to intervene within the sight test market since the sight test is an important tool by which certain eye diseases

can be detected thus reducing the potential for future cost implications to the health service. Effective policy is required to determine whether it is important to provide a direct subsidy that provides certain groups with free sight testing or provide the entire community with free sight tests, in other words whether it is best to have targeted or comprehensive provision. In order for there to be appropriate and effective policy development it is necessary to establish the factors that would influence sight test attendance be they individual's social background or resources.

As we have said before the sight test is a two pronged tool, with both screening and health improvement aspects, the screening aspect is of interest to the government (potentially reduce future costs) whilst the health improvement aspect would be of interest to individuals (sight improvement by use of glasses). Although this paper examines the sight test this health service can be used as a proxy for other similar health services that provide a shorter term health improvement as well as a longer term screening benefit.

In this paper we first examine the sight test as a tool and the possible direct health improvements and screening benefits offered. The next section examines the economic theory of health improvement and applications to the sight test. The final section examines the economic theory of screening and the application of the sight test to the screening argument.

Any empirical analysis carried out is reported in a subsequent paper and examines the individual's decision to attend for a sight test based upon the

human capital approach, where the sight test is seen as an investment towards the individual's human capital stock.

2.2 *The Sight Test*

2.2.1 What are the direct health improvement benefits of the sight test

It is important to highlight the sight test's attributes and components before we apply economic theory to this good so that we fully appreciate all aspects of the sight test. In this section we examine the health improvement component of the sight test.

The first and possibly most obvious output of the sight test is the potential improvement in vision. The sight test can detect problems such as short or long sightedness among others (Colleg (1998)). These informational outputs, regarding the level of vision, when utilised appropriately by individuals can lead to improvements in sight and quality of life through vision correcting aids, such as glasses or contact lenses.

In this respect it could be said that sight tests and spectacles are linked to one another. A prescription for glasses is a possible complementary good to sight testing. The theory of the sight test and glasses link is not as clear as that of other complementary goods such as, for example, the theory surrounding CDs and CD players. One can attend for a sight test and not require a prescription for glasses, one may also attend for a sight test and obtain a prescription for glasses, which is or is not dispensed; on the other hand it is possible to purchase a pair of 'reading' glasses over the counter without a prescription or even attending for a sight test.

2.2.2 Health Improvement literature

Applying simple demand theory to sight tests would imply that as the cost of sight testing falls to zero so demand would increase; however there are other factors that need to be taken into account, specifically the risk of requiring glasses.

If individuals are not required to purchase glasses and they are eligible to receive free sight tests (thus the price of the test is equal to zero) individuals may still not, rationally, attend for a sight test, since the risk of requiring glasses may increase the shadow price of the sight test to more than zero. If sight tests are free to the over 60s, then it does not automatically follow that there will be an increase in uptake. The over 60s also face the risk of requiring glasses, the price of which may not equal zero. Even if they are able to claim assistance with the cost of glasses, they may not seek this assistance, this can be assumed from RNIB campaign document for better health in Wales, where approximately 20% of individuals with diabetes, glaucoma or a first relative with glaucoma did not know they were eligible for a free sight test[4].

The sight test is linked to risk of requiring glasses, which inevitably increases the shadow price of the sight test, the available literature does not examine a direct link; we have aimed to provide indirect evidence of the possible implied link that may exist between these two health care interventions.

[4] RNIB (2001) section 5.3

Smeeth and Iliffe (1998) examine the effect of the possible introduction of a screening programme for the over 65's and postulate that the risk of requiring glasses is an important factor in individual attendance for a sight test. This systematic review of randomised trials did not find any trials that were primarily assessing visual screening. Smeeth and Iliffe believe that there are great barriers to sight test attendance, namely that the fear of costs is a significant barrier to the attending the optometrist and obtaining glasses.

Reinstein (1993) examined the correctable undetected visual deficit in patients over 65 attending an accident and emergency department and found that half the patients found to have a visual acuity problem, that was undetected and correctable, had not attended for a sight test for 2 years mainly due to financial considerations.

2.2.2.1 Summary of the sight test as a health improvement tool

We believe that the demand for the sight test is linked to the price of glasses, and the risk of requiring them, for the over 60 population who may not be, or may not know that they are, eligible for assistance with the cost of glasses. The link seems evident with both Reinstein (1993) and Smeeth and Illiffe (1998) who highlight the link between sight tests and the financial considerations that come with it.

2.2.3 Historical and eligibility perspective of the sight test

If the sight test does have direct health benefits which sets it apart from other screening tools the impact of price upon the uptake of the sight test needs to be assessed.

Until April 1989[5] the sight test was available free to all individuals on the NHS, however spectacles and contact lenses had to be paid for, unless the individual was eligible for government assistance.

Since April 1989 up to the present time all individuals have to pay for their sight test unless they fall into one of the eligibility categories outlined below.

The eligibility of an individual for a free sight test is restricted to the following groups: all children under the age of 16, all students in full-time education aged between 16-18, individuals' over 60, adults receiving income support and their partners, adults on family credit and their partners, people on low incomes that have an AG2 form and their partners, all glaucoma and diabetes sufferers, individuals registered partially blind or partially sighted, people requiring complex lenses, close relatives aged 40+ of a glaucoma sufferer and adults receiving disability workers allowance. Those individuals that are housebound and fall into to one of the eligibility groups above are entitled to a visit by an ophthalmic practitioner at home, this is known as a domiciliary visit[6].

[5] FODO (1996) (p.1)
[6] FODO (1997) (p.2)

Before July 1986, individuals eligible for government assistance for spectacles were eligible only for NHS spectacles. After July 1986[7], the government introduced a voucher scheme which provided individuals with vouchers to put towards the cost of their glasses. The eligibility criteria are as outlined below. At first, this eligibility scheme only operated for the purchase of spectacles however, it was later, in April 1988[8], extended to the purchase of contact lenses.

The eligibility of individuals for voucher assistance towards spectacles or contact lenses is restricted to the following groups of individuals: all children under the age of 16, students in full time education aged between 16-18, adults on income support and their partners, adults on family credit and their partners, adults on disability workers allowance and, individuals with low incomes in possession of an AG2 or AG3 form and their partners[9].

Also, in April 1989[10] the sale of ready-made reading glasses was extended to unregistered suppliers, so that individuals could purchase a pair of reading glasses 'over the counter' without having to go to an optician for a sight test.

Table 1, below depicts the number of sight tests and vouchers (for spectacles) paid for by Health Authorities for each year in England. This data represents sight tests and vouchers available on the NHS from 1989/90, when sight test fees and eligibility criteria for free NHS sight tests was introduced to year

[7] Department of Health (1996) (p.1)
[8] Department of Health (1996) (p.4)
[9] FODO (1997) (p.2)
[10] Department of Health (1996) (p.1)

2000/1 when the over 60s became eligible. From 1989/90 to 1990/1 we can see a decline in the number of NHS sight tests, after which there is a steady increase to 6.99 million in 1997/8 where it remains constant until 1998/9. In 1999/2000 we see an increase to 9.40 million when the over 60s are included in the eligibility criteria. With the NHS vouchers scheme we see a steady increase in vouchers dispensed from 2.27 in 1989/90 to 3.97 million in 1996/7, after which there is a decline to 3.58 million in 2000/01.

Table 1: Number of sight tests and vouchers paid for by HAs (FHSAs before 1996-7): by year[11]

England		
Financial Year	**Sight Tests** (millions)	**Vouchers** (millions)
1989-90	5.28	2.27
1990-1	4.15	2.43
1991-2	4.98	2.84
1992-3	5.53	3.19
1993-4	5.93	3.48
1994-5	6.38	3.74
1995-6	6.51	3.82
1996-7	6.81	3.97
1997-8	6.99	3.94
1998-9	6.99	3.78
1999-2000	9.40	3.66
2000-01	9.57	3.58

This data reports the NHS sight test uptake; there is no data available to assess the sight test attendance for ineligible groups.

[11] Department of Health (2001)

2.2.4 What are the screening benefits of the sight test

In this section we examine the sight test and its attributes as a viable screening tool, and how it can be applied to the preventative health care market to reduce potential future costs.

The purpose of screening in medical care is to act as a preventative measure against various health problems. There are two forms of intervention: primary and secondary. Primary preventative intervention is the form that will stop the health problem occurring, such as immunisations. Secondary preventative intervention will attempt to 'capture' the problem as early as possible so that treatment can be administered quickly to gain the most advantage and act against the disease; this is where screening fits in.

Screening is an important form of detecting disease in its early stages and therefore providing treatment that can reduce the likelihood of the disease progressing further, as well as reducing the amount of discomfort and suffering. The main detection form is to use a simple diagnostic test, ideally before the individual being screened starts to suffer from any symptoms.

There are many various methods of screening. Mass screening is the large-scale screening of whole population groups, selective screening describes screening of certain high-risk groups in a population, multi-phasic screening is administration of two or more screening tests to large groups of people, surveillance implies the long-term assessment of individuals, and case-finding is the screening of individuals already in contact with health services

for purposes of detecting disease and bringing patients in for treatment. There is an important difference between routine screening, where individuals are invited to be screened, and 'opportunistic' screening, where the patient is the one initiating contact with the health service, at which point screening services may then be suggested to them.

The type of screening proposed by using the NHS provision of the sight test to the over 60s is selective opportunistic screening able to detect glaucoma, cataract, age-related macular degeneration (ARMD) and diabetic retinopathy (DR).

One of these is primary open angle glaucoma (referred to as glaucoma in this paper), this is a problem that affects the vision of the individual to such an extent that if left untreated will cause permanent irreversible damage and could lead to blindness. There are of course many other vision problems that the sight test can detect, these being cataract, diabetic retinopathy (DR), age related macular degeneration (ARMD) and problems relating to incorrect fitting of glasses or contact lens.

Cataracts can cause sight problems, individuals may not be able to see clearly, their vision may become blurred but cataracts are operable. Diabetic retinopathy (DR) affects diabetic patients where they will experience black spots in their vision and slowly their vision if left untreated will be lost. Age related macular degeneration (ARMD) can not be cured but can be managed so not causing further damage or sight loss. Glaucoma is caused by

increasing pressure within the eye, causing the visual field to decrease and so leading to permanent loss of sight, this disease is manageable, limiting the potential loss of sight.

To assess the extent of the visual problems potentially affecting the pensionable population, there was a study conducted upon a representative sample of persons of pensionable age (65 years old) and over, the sample was taken from North London. The North London Eye Study (NLES)[12], aimed to identify the prevalence of certain eye diseases in these individuals; the prevalences detected are as below:

Table 2: Average prevalence figures for eye problems for the over 65s.

Eye problems	Prevalence[13]
Cataract (causing visual impairment*)	30%
ARMD (causing visual impairment*)	8%
Glaucoma (definite cases)	3%
Glaucoma (suspect cases)	7%
Refractive Error (causing visual impairment*)	9%

* Vision in one or both eyes is less that 6/12, the legal requirement for a driving licence.

Of the above problems that were detected, only cataract and refractive error can be corrected. The other two eye diseases, ARMD and glaucoma, if left untreated can lead to irreversible sight loss, leading to a substantial number of individuals over 65 (~18%), in a state of unnecessary visual impairment.

[12] Reidy, et al. (1998)
[13] Reidy, et al. (1998) (p1644)

The NLES study found that quite a high number of individuals with visual problems were not even in-touch with eye services (Table 3), indicating the possible extent of the problem, individuals are going through life with eye problems that could be dealt with.

Table 3: People with vision problems, not in touch with eye care services

Eye Problems	Not in touch with eye care services[14]
Cataract (causing visual impairment)	88%
ARMD (causing visual impairment)	86%
Refractive Error (causing visual impairment)	96%**
Glaucoma (definite cases)	74%
Glaucoma (suspect cases)	84%

** 30% of this figure had visited an optometrist in the past 12 months.

According to the table above the sight test could potentially detect and rectify the vision of a significant proportion of the over 65 age group suffering from some sort of visual impairment.

A number of epidemiological equations have been calculated in order to establish the possible number of cases of glaucoma in a given population. These equations are known as Quigley- Vitale (QV) and the Tuck-Crick (TC) epidemiological equations. Both of these equations assess the prevalence of glaucoma based upon a number of studies carried out in the US (QV) and Ireland, Italy, and Australia (TC). These equations are very sound and have

[14] Reidy, et al. (1998) (p1645)

good predictive prevalence values; they do not include the NLES prevalence data, which is important because it is based in the UK.

The table below (Table 4) shows the overall prevalence rates of glaucoma, as calculated by the QV and TC equations and, NLES data, which excludes the Afro-Caribbean ethnic group; this is because the prevalence rates for this group are much greater.

Table 4: Prevalence of glaucoma as calculated by the TC and QV equations, as compared to the actual prevalence seen in the NLES (which excluded the Afro-Caribbean ethnic group)

Calculation Method	Glaucoma Prevalence
Actual prevalence in the NLES sample	3.01%
TC calculated prevalence	2.73%
QV calculated prevalence	3.50%

The next step is to estimate the potential number of glaucoma sufferers that could be detected, based upon the prevalence of glaucoma and data on age structure and ethnic mix in England and Wales. It is important to highlight the fact that the population prevalence estimates were calculated excluding the Afro-Caribbean ethnic group because; the incidence of glaucoma in this ethnic group is much higher.

Table 5: Glaucoma prevalence in the over 60's population of England and Wales (not including the Afro-Caribbean ethnic groups), the 3 estimation methods were used; these are the TC, the QV prevalence equations, and the prevalence data from the North London Eye Study.

Age	E & W Population	TC Equation		QV Equation		Estimates from NLES	
		Number	%	Number	%	Number	% (SE)
60-64	2,525,683	24,346	1.0	37,019	1.5	--	--
65-69	2,471,923	37,343	1.5	51,470	2.1	39,030	1.6
70-74	2,005,106	44,971	2.2	57,510	2.9	48,433	2.4
75-79	1,658,753	51,593	3.1	63,426	3.8	40,212	2.4
80-84	1,111,272	44,462	4.0	54,993	4.9	47,088	4.2
85 +	761,651	39,005	5.1	52,965	7.0	53,764	7.1
All	10,534,388	241,721	2.3	317,383	3.0	228,526	2.9 (0.4)

Abv: E&W – England & Wales; TC – Tuck-Crick; QV – Quigley-Vitale; NLES – North London Eye Study; SE – Standard Error

In Table 5 above, the prevalence data for the 60-64 age group is not available from the NLES; this is because the survey only had a sample of individuals over the age of 65. However, as can be seen from Table 4, prevalence estimates provided by QV and TC are very close to those provided by the NLES data, the TC estimates are especially similar.

The NLES found that, currently, the detection rate of definite glaucoma cases is 1 in 3[15], however it is generally accepted that the detection rate of definite glaucoma cases is 1 in 2; these detection rates are based on patients in touch with the eye services (optometrists).

[15] Reidy, et al. (1998) (p.1645)

Therefore, given that the TC and QV prevalence figures for glaucoma are used, since the NLES study cannot given prevalence for the 60-64 age group, the potential for glaucoma detection is huge.

Table 6: Detection rates of glaucoma in the over 60s, in England and Wales (excluding the Afro-Caribbean ethnic group)

Total No. Definite Glaucoma Cases and Detection Rates					
Prevalence	Detection Rate		Prevalence	Detection Rate	
TC used	1 in 3[1]	1 in 2[2]	QV used	1 in 3[1]	1 in 2[2]
241,721	80,573	120,860	317,383	105,794	158,691

[1] Detection rate found by the NLES

[2] Detection rate generally used.

Thus, from Table 6 it can be seen that the potential for increased detection of definite cases of glaucoma is between 80,000 and 150,000, if the sight test is provided free to the over 60s, and there is a 100% take up rate.

Based upon the data from the NLES paper, we can propose that the sight test is a valid opportunistic screening tool for various important eye health problems such as cataract, glaucoma and, so forth. The sight test is a valid tool by which potentially damaging vision problems can be detected earlier thus allowing the individual and society the opportunity to limit future costs.

2.2.5 Applying the sight test to the screening perspective

Having previously examined the applicability of the sight test as a potential screening tool this section applies the sight test to a screening perspective to examine the validity of this.

Screening as a process or health program is very closely examined by Holland and Stewart (1990). They identify screening as the process of actively seeking to identify a disease or pre-disease condition in asymptomatic individuals. The aim of screening is to detect disease before the symptoms are present and an individual decides to seek medical advice.

Screening via the sight test has the aim of protecting the eye health of individuals whilst also detecting potential future problems in order to reduce future costs and increase benefits. The earlier the disease is detected so the sooner the treatment can begin thus increasing the chance of preventing irreversible sight deterioration and sustaining a higher quality of life.

Screening not only has two outcomes, one a clear benefit and the other no effect, but it can also cause harm. Inevitably the benefits of screening are: improved diagnosis for some cases detected by screening, less radical treatment may be needed to cure early cases, resource savings, and the obvious reassurance to those with negative results. However, in the opposite vein, disadvantages of screening include: a longer morbidity for untreatable detected cases, over treatment of abnormalities that are questionable, resource

costs, false reassurance to false negative individuals, anxiety and, morbidity for those with false positive results and the possible hazard of screening tests.

Before, recommending a screening programme as possible policy, a proper assessment of the resource implications, versus benefit, is required. Even though Holland and Stewart recommend a thorough assessment of resource implications, it may not be necessary in the case of the sight test as a screening tool, since the test is not recommended as a screening programme but as an opportunistic screening program tool.

Holland and Stewart (1990)[16] set out ten basic principles of screening:

The condition sought should be an important health problem

There should be an accepted treatment available for patients with the recognised disease.

The facilities for diagnosis and treatment should be available

There should be a recognisable latent or early symptomatic stage

There should be a suitable test or examination

The test should be acceptable to the population

The natural history of the disease, including the latent period until declared disease, should be adequately understood

There should be an agreed policy on whom to treat as patients

The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.

[16] Holland and Stewart (1990b) (p.12)

Case finding should be a continuing process and not a 'once-for-all' project.

With respect to glaucoma detection via the sight test, it would be pertinent to point out that all the factors above are significant. Glaucoma is a significant health problem, which is validated through the evidence provided earlier on in the paper. Once patients are recognised, or are suspected, as having glaucoma they are referred to an ophthalmologist, who will follow-up with treatment for their condition. Glaucoma is a disease that is asymptomatic; it can not be detected by the individual concerned but is detectable by an ophthalmologist. The test for glaucoma involves testing intra-ocular pressure and visual discs, both these tests are available at ophthalmology departments and have been used widely. Research in glaucoma, its prevalence, disease pathway, detection and treatment methods are well researched. In terms of the cost of case-finding in glaucoma, there is no set screening programme for glaucoma, however the detection of glaucoma can be greatly increased by individuals presenting for sight tests, and those that are most at risk of developing glaucoma being tested for this. This will inevitably mean that the more individuals that are detected, so the more individuals that can be treated before their vision is irreversibly lost. In terms of case-finding being a continuous project, individuals are encouraged to attend for a sight test at regular intervals therefore individuals will have a greater chance of continuous detection. Free sight testing for the over 60s, in theory, means there will be an increased likelihood of sight test attendance and subsequent detection.

Holland and Stewart (1990)[17] examine the cost-effectiveness of various screening programmes that an individual may be involved in during their lifetime, from childhood through to adulthood and old age. However, since this thesis is in essence only looking at the effect of screening for sight problems, such as glaucoma, cataract and so forth, in the over 60's, via the sight test, it will seem relevant to look at only screening programmes, which Holland and Stewart (1990) have dedicated to the old age section of their work.

There are two methods of screening elderly individuals; the first method is through the selection of high-risk groups, the second through opportunistic screening, much like the sight test. Holland and Stewart (1990) indicate that the aim of screening is not to offer a cure from the diseases suffered by the elderly but, to offer treatment in order to improve their quality of life. Holland and Stewart (1990) do examine the aspect of vision screening in the elderly. Visual problems can lead to problems with walking and functions of daily life, they agree that many elderly people may already have reading glasses however; a change in prescription is required as age increases[18]. At the time of publication of their work Holland and Stewart rightly claim that there is no national policy on the screening of glaucoma, and this is still presently correct. Public awareness of this condition is generally low, and many general practitioners are not aware of the difficulties that it can cause, however it is still an important cause of blindness in the UK[19]. Glaucoma is a complex condition, which has many contributing factors. There is not a

[17] Holland and Stewart (1990b) (p.14)
[18] Holland and Stewart (1990b) (p.202)
[19] Holland and Stewart (1990b) (p.202)

simple universal screening test, and the only treatment available in 1990 is to lower intra-ocular pressure. It had been suggested that screening for glaucoma be part of the routine eye examination, as ophthalmologists and optometrists need to be aware of their influence on the detection of glaucoma. Holland and Stewart (1990) found that most referrals for glaucoma came through general practitioners from optometrists. It seems clear from the brief synopsis carried out by Holland and Stewart, that attention is warranted for the problems of vision screening in the elderly. Holland and Stewart suggest that to encourage regular testing, the over 65 population should be added to the list of individuals exempt from payment Holland and Stewart (1990), this has now occurred (Daily Telegraph Oct. 12 (1999)).

Thornton (1999) is an advocate of private screening and believes in the limited and questionable benefits, as well as potential harm, to be achieved by offering extensive public-funded adult screening. Thornton postulates that whether an individual attends for a specific screening programme, is a matter of choice and there does not seem to be any objective or scientific criteria that could be applied, by doctors, to enlist individuals into a screening programme.

Thornton (1999) looks at various screening programmes currently offered and tries to discuss which of these programmes should be state funded. One such programme is vision screening. At the time of Thornton's work, all adult vision testing was private unless, one was eligible for an exemption. To be exempt from paying for a sight test the individual would need to fit into

one of the following categories: all children under the age of 16, all students in full-time education aged between 16-18, adults receiving income support and their partners, adults on family credit and their partners, people on low incomes that have an AG2 form and their partners, all glaucoma and diabetes sufferers, individuals registered partially blind or partially sighted, people requiring complex lenses, close relatives aged 40+ of a glaucoma sufferer and adults receiving disability workers allowance.

Thornton recommends that the NHS should not screen for vision problems, since most refractive disorders and cataract can be treated when vision becomes a problem. The only exception to this point is glaucoma, where symptoms will occur for a period of time before blindness occurs, this disease can be detected by optometrists through visual field or intra-ocular pressure measurement.

Thornton uses the argument that since, there has been a fall in the number of people attending for eye tests after the introduction of the test fee, then people have presumably indicated their preference to consume other goods, and so sight testing is not highly valued, and should not be placed under NHS provision.

Thornton's argument that given a fall in sight test after the introduction of fees and eligibility groups we should assume the sight test is not very highly placed in people's preference ranking is flawed. Firstly, the fall in sight testing after the introduction of fees and eligibility in April 1989 simply

indicates that there has been a fall in the uptake of free sight tests, however this trend can not be declared as necessarily indicative of the market as a whole. Even if we were to accept that the decrease in sight testing post 1989 is indicative of the entire market, we can not use this as an argument to not provide the over 60s with free sight testing. Secondly, glaucoma is a debilitating disease, which if left alone will cause irreversible sight damage and blindness. The over 60s have an increased risk of suffering from this disease without knowing it, given that the disease is asymptomatic and can be mistaken as part of the 'growing old' process.

In economic terms individuals will maximise their utility subject to a budget constraint, it is well appreciated that once individuals are over 60 they retire and face a much decreased budget constraint, it is subject to this constraint that individuals will try to maximise their utility. The sight test in itself is not necessarily a costly good, however if this is weighed against other goods that are now competing for the same budget it may well fall by the way-side. Therefore, though Thornton may be fully correct in assuming that if individuals are not purchasing a sight test now that they have to pay for it they do not prefer it highly, this argument should not be used against the over 60s. This age group is more at risk of glaucoma, than those under 60, and has a lower income level, thus will potentially benefit more from a free sight test, there is also a greater benefit to society.

2.2.6 Literature review of the sight test as a screening tool

In this section we review the literature in vision screening that has been published in order to evaluate whether this answers our question regarding sight test as a viable screening option, especially with relevance to glaucoma.

Glaucoma screening has been examined in a number of published papers, which are discussed below. A modelling study of the cost-effectiveness of glaucoma screening was carried out by Gottlieb, et al. (1983) concluding that a targeted rather than mass glaucoma screening programme would be most efficient method of using scarce health resources. The authors designed a model of the process of glaucoma screening, the diagnosis and, treatment, in order to assess the cost of screening in relation to quality-adjusted years of vision saved. A number of assumptions were made in the model regarding the accuracy of the tests, the natural history of the disease, and the effectiveness of the treatment methods. Gottlieb and colleagues found that glaucoma screening, when targeted at certain groups of the population, is probably cost-effective. They found that the screening of select groups (that is: Afro-Caribbean population, diabetic population and relatives of glaucoma patients) is probably more cost-effective than the screening of the general population. It would be cost-beneficial if efforts were directed more towards improving follow-up and compliance rather than actual screening efforts.

A prospective survey carried out by Tuck and Crick (1991) covering 5% of all sight tests performed in England and Wales by optometrists over a six month period showed that the sight test was a useful tool with which to detect

vision problems; however greater co-operation is required within health professionals before the benefits of the sight test, in terms of disease detection, can be realised. Using study data the efficiency of referral for suspected glaucoma to general practitioners and consultants by optometrists was examined. The trial reports that 0.9% of individuals aged over 40 attending for a sight test were referred with suspected glaucoma; of these 90% attended for an examination by a consultant ophthalmologist, 41% were confirmed as having glaucoma, and a further 32% required monitoring. The survey found that greater co-operation is required between optometrists, ophthalmologists, and GPs, as well as greater encouragement of optometrists to use all three main tests (ophthalmoscopy, tonometry and perimetry) in patients before referral.

Following on from their survey Tuck and Crick (1997) conducted a study of the most appropriate method of screening for glaucoma and found that screening for glaucoma is most likely to be economic when conducted in conjunction with overall eye examinations. The study used ophthalmoscopy, tonometry, and perimetry in different combinations to assess which method would be cost-effective (where the additional benefit obtained by implementing the new intervention is judged to be worth the additional cost of the new intervention – this is based upon an implicit cost-effectiveness level). The study population used were white Caucasian persons over the age of 40, and was carried out in the UK (London). Tuck and Crick used epidemiological data regarding the aetiology of the disease and sensitivity/specificity of the screening modes extracted from literature

between 1966 and 1995. Cost data and resource use data was collected from literature between 1983 and 1994. The study prevalence of primary open angle glaucoma (POAG) was 1.2% in the over 40 white Caucasian population of London (UK). Based on the literature an assumption was made that undetected POAGs would be half of all POAGs. Tuck and Crick concluded that glaucoma screening of individuals over the age of 40 could be cost effective and thus justifiable, if the screening is worth more than $850 to detect a new case. The modes of glaucoma screening found to be most cost effective in detecting cases are methods using a combination of ophthalmoscopy and tonometry routinely on patients over 40, with perimetry routinely on high risk groups.

Diabetic retinopathy is also a sight threatening disease that can be managed with appropriate treatment and limit potential sight loss if picked up earlier. Bachmann and Nelson (1998) conducted a modelling study to assess the impact of diabetic retinopathy screening on a British district population, and found that screening for diabetic retinopathy is a cost-effective methodology. The objective of the study was to develop a simple model through which it would be possible to estimate the probable results from a screening programme and its impact upon the incidence of diabetic retinopathy in a diabetic population. Screening for diabetic retinopathy was found to be an easy and effective method, which could be improved by improving current test performance and increasing uptake amongst the most 'at risk' diabetics.

There are various eye diseases that can be assessed via screening (using different methodologies) including a community based screening program; Smeeth and Iliffe (2001) carried out a Cochrane-based systematic review on community screening for visual impairment in the elderly and found that community-based screening of asymptomatic older individuals did not necessarily result in vision improvements. The review looked at five multi-component trials, providing visual outcome data for approximately 3,500 people aged 65 and over; length of follow-up was between two and four years, and self-reporting was used for visual impairment. This review found that there was no evidence to suggest that community-based screening of asymptomatic older people resulted in vision improvements. However, the lack of effectiveness is explained through a number of factors; visual assessment was part of a component screening package so results may have diminished, and it may have been more effective if visual impairment for screening had been carried out in isolation. Secondly the review looks at the effectiveness of visual impairment screening alone however; screening on its own will not improve vision; vision improvements will need interventions to follow screening; therefore it is possible that the trials in this review did not adequately follow-up vision problems detected. Thirdly, the lack of effectiveness could have been a cause of individuals not acting upon advice given to seek further care; it may be that participants in these trials did not feel the need to go for further treatment or may have felt that the barriers to further treatment were too great, such as the cost of further eye testing, cost of glasses and so on.

Goddard (1990) examined what impact the removal of the screening part of the sight test would have upon welfare and found that the removal of the 'screening' part of the sight test would be detrimental to societal and patient welfare. In this study Goddard looks at whether there will be an effect upon resource use and quality of patients' life if the glaucoma and diabetic retinopathy screening element is removed from the sight test as it stands. Goddard examines the glaucoma and diabetic retinopathy incidence and prevalence literature to establish whether de-regulation of the sight test to a two-tier examination, where patients have a refractive eye exam and then can request a fuller eye exam, will have an effect upon detection rates and quality of life. The conclusion is that the effect of such a two-tier system would be detrimental, since many glaucoma sufferers would be asymptomatic and so would not necessarily request a fuller eye exam. Also, diabetic retinopathy sufferers would not be as closely monitored. This paper does not look at the effect of introducing sight test fees, but looks at a separation of the eye exam into a refractive error exam and a fuller eye exam that detects disease. This paper only looks at the impact of a two-tier system upon glaucoma and diabetic retinopathy, ignoring the essential problem of cataracts.

The sight test, as alluded to before, can pick up vision problems as well as eye diseases thus sight test attendance and the impact of factors upon this are important. An unpublished paper, Sim and Vafidis (1990), examining the effects of the April 1989 introduction of sight test fees upon patients and services in the Barnet Health Authority, found that the number of sight tests carried out by opticians had reduced, with the effect being more apparent in

the elderly population. The report also found that in the Barnet (London, UK) area there was an increase in the number of elderly attendees at hospital outpatient clinics, though no direct evidence of inappropriate referrals was found. Although the authors did not find any evidence of a decrease in the incidence of glaucoma or the number of outpatient referrals not followed up this could be because the study was carried out when the sight test fee introduction was in its infancy, so the effects upon outpatients may not have been fully realised. This paper looks at the opposite of what we wish to examine. They examine the effect of introducing charges, whereas we would like to examine the effect of removing charges upon uptake of sight testing by the elderly. This paper, if taken as it stands, can be used to claim that since the introduction of the sight test has had a greater detrimental effect upon the elderly uptake of sight testing then, removing the fees for this group would increase uptake thus, reversing any inappropriate use of outpatient clinics that may be occurring due to the sight test fee introduction.

2.2.6.1 Summary of the sight test as a screening tool

Reidy et al (1998) in the NLES show that there is a potential for increased detection of definite cases of glaucoma to the tune of between over 80,000 and over 150,000, if the sight test is provided for the over 60s free, and there is a 100% take-up rate.

A review of the eye disease screening literature shows that screening for glaucoma is found to be both effective and cost-effective (Tuck and Crick 1997, Gottlieb et al 1983), screening for diabetic retinopathy is found to be effective (Bachmann and Nelson 1998). Goddard (1990) found that there

would be a detrimental effect upon quality of life and pick-up rates of glaucoma and diabetic retinopathy if the two-tier system is introduced suggesting that the current sight test is a valid screening tool.

The literature shows that glaucoma is a significant vision problem, however the literature surrounding whether a national screening programme being set-up is not clear cut. Given that a national screening program is not in place in the UK, and we are not trying to suggest that one should be, the opportunistic screening capabilities of the sight test are very valuable, and ought to be examined closely.

2.3 *Demand for Health Improvement*

2.3.1 Economic theory of health improvement

Before examining the sight test as a health improvement tool it is necessary to examine the demand theory surrounding the general health improvement services.

The demand for health (or health improvement) can be used to derive the demand for health care, which is influenced by the consumer's willingness and ability to pay for the fulfilment of their desire or need.

The demand for health care is a better measure than the demand for health. All individuals demand better health in order to maximise utility, however this demand can not readily be translated into a measurable form. Demand for health care, on the other hand, can be measured quite simply as the number of people that, or the number of times a person will, go to the optician, dentist, GP or some other health care provider. When an individual is attending for a sight test, s/he is depicting demand for better vision as a consumption good and as an investment good to allow them to undertake other activities and participate in the labour market freely.

The following section examines very closely the Grossman model and the applicability to the sight test. Grossman's model uses a human capital framework model and makes a major contribution to the understanding of the

demand for health care (proxy for demand for health); which is very relevant to demand for the sight test.

2.3.2 Grossman's model

Grossman (1972)[20] developed a theory of demand for health care based on the premise that health is a commodity produced by individuals using health care and their own time as inputs, thus the demand for health care is derived from the demand for health.

In Grossman's model the demand for health is subject to factors other than price of health care, individual's income or preferences.

Grossman's model is based on the human capital framework, the human capital framework model assumes that individuals inherit a stock of wealth, which can be added to (investment) or taken away from (consumed). This concept is applied to the health capital framework, which is a component of human capital. Individuals inherit stock of health which can be consumed or invested in; this initial stock of health will begin to depreciate over time (after a given point) until the health stock reaches such a minimum point as to cause death. To maintain or increase the initial stock of health, or reduce the depreciation rate, the individual must invest in his/her health stock.

[20] Grossman (1972)

2.3.3 His assumptions

Grossman's model assumes that individuals are making efficient choices without any outside interference – this assumes that individual's are fully aware of their initial health stock, the rate of depreciation they face, have full knowledge of the method by which to produce more health and what their optimal health level is.

Grossman assumes that the individual is a utility maximising consumer, where utility is a function of healthy days (h_i) in the specific period and the consumption of other commodities (z_i). Therefore, good health (H_i) is an important element of the individual's preference function determining the ease with which market and non-market activities can be carried out. There are two reasons why people want to have good health, an increase in healthy time means an increase in productivity and, consumption during healthy time will derive a higher utility.

The individual's length of life is endogenous to the model, when the level of health stock (H_i) reaches a minimum (H_{min}) death will occur. The initial 'inherited' stock of health is an exogenous factor.

Grossman assumes that 'good health' is a durable commodity, which means that health will last for more than one period. Good health is assumed to be both an investment and a consumption commodity.

In his model Grossman assumes that the individual has a total sum of time available in any period (Ω_i) which consists of time spent on earning income (hours of work) (TW_i), time spent producing gross investment in health (TH_i), time spent producing composite commodity (T_i) and time lost due to illness (TL_i):

$$\Omega_i = TW_i + TH_i + T_i + TL_i$$

This means that of their fixed time budget individuals have to spend their entire time earning an income, investing in health, producing 'other' commodities or being ill. TL_i is inversely related to H_i, this means that if we assume h_i (the total number of healthy days produced by H_i) to be the total number of healthy days in a given period then $TL_i = \Omega_i - h_i$. Individuals can only earn income or produce 'other' commodities if they are well therefore sick time in any period is equal to the total time in period 'i' minus the healthy days in period 'i'. However, the individual's time spent on gross health investment (I) (such as GP regular check-up) is separate from sick time (GP visits due to illness).

For an individual to increase his stock of 'good health' in period 'i' then the individual needs to increase gross investment in health in the period before (i-1). An increase in the individual's time spent investing in health (if depreciation is constant) in period 'i' would mean an increase in I_i and H_{i+1} and a decline in TL_{i+1}.

To maintain his/her level of health stock the level of net investment made by the individual in period 'i' is equal to the gross investment in period 'i' minus the depreciation, where the depreciation rate is an exogenous factor though is dependent upon the individual's age.

$$H_{i+1} - H_i = I_i - \delta_i H_i$$

2.3.4 How does the model fit together

So the individual demands health for consumption (to yield direct utility) and investment (health increases the time available for market and non-market activities). Both consumption and investment aspects of the model are linked, individuals will invest in health to increase (or maintain) their healthy time, which means they have a greater time available on market and non-market activities (less time being ill) and thus derive a greater utility.

The optimal investment (demand for health) for the individual in Grossman's model is when the marginal cost (of investing in health) is equal to the marginal benefit; this is defined as the marginal efficiency of capital (MEC):

$$\gamma_i = r + \delta_i$$

where:

r = Rate of interest on other investments;

δ_i = Rate of depreciation in period 'i' of health;

$\gamma_i = W_i G_i / \pi_{i-1}$; where W_i = Wage rate in period i, G_i = Marginal product (rate of return) of health capital in period 'i' which diminishes due to diminishing

marginal returns, π_{i-1} = marginal cost of gross investment in period 'i-1' – the direct cost of investment in health;

Grossman's model incorporates both consumption and investment aspects of health demand; the consumption model suggests a positive correlation between the rate of depreciation and age, which generates a stock of health that is consistent with a finite life.

Above is the investment model, where the optimal amount of gross investment in health would be found by equating the marginal rate of return on an investment in health (W_iG_i/π_{i-1}) to the cost of health capital ($r + \delta_i$).

Demand for health is dependent upon the shadow price of health which is affected by age, income and, education. The shadow price of health increases as age increases, then the depreciation rate of health increases over the individual's life-cycle, the marginal cost of health capital increases and so the demand for health capital will fall, however there could be an increase in demand for health care (the life-cycle effect). Education increases the marginal efficiency of capital, which means that the more educated an individual so the more efficient a producer of non-market goods (marginal product is increased) which increases the optimal health stock; increased education decreases the shadow price. An increase in wage increases the return on healthy days so the level of optimal health stock is much higher, however investment in health care also requires time which increases the cost of health investment (ie greater opportunity cost).

Grossman's empirical work[21], uses data from the 1963 NORC (National Opinion Research Centre) sample of the US National Health Survey encompassing 2,637 families or 7,803 persons. The paper shows that education has a positive and significant effect upon demand for health, increasing age in turn leads to a decrease in health and an increase in medical expenditure, the wage elasticity of health is positive and significant indicating that an increase in wage (earned) will lead to an increase in the individual's health whilst an increase in income (unearned) will lead to a decline in health since healthy time is negatively income elastic.

2.3.5 What are the implications for the demand and price of health services based upon Grossman's model

Grossman's model implies that if education level were to increase individuals would be more efficient investors of health, similarly if individuals (who are low-paid) earned a higher income they would also be more efficient health investors. If the price of health care was reduced, especially where individuals were not earning a high income they would be a greater investment in health stock (lower cost of health care means individual can invest more in health stock with the same budget, so increasing his/her health). If the marginal cost can be equated to the marginal benefit of health capital the individual will maximise his/her health stock.

[21] Grossman (1972) (p.xvi)

2.3.6 How does Grossman's model differ from the other literature

In usual economic theory, and within Grossman's model, the consumer is assumed to be a rational consumer able to make utility maximising decisions based on a fixed budget allocation. Cullis and West (1979)[22] , however, argue that this view of rational utility maximisation is inappropriate for the health care market. The consumer's level of knowledge regarding his state of health and the consequences of undergoing certain treatments will probably be different under different assumptions, thus leading to a divergence from theory. One extreme assumes that the consumer has all the knowledge required to make a decision (perfect knowledge) the other extreme assumes that the consumer knows nothing (asymmetry of information towards the consumer). Grossman's model assumes that the individual has perfect knowledge and so will be able to maximise his/her health stock by consuming health care (investing in health stock).

Fuchs (1972[23]) looks at the growing demand for medical care, and is one of the first of such papers that tries to examine the empirical evidence available and explain the phenomenon occurring in the medical care field with the use of economic tools. Fuchs (1972) highlights the fact that demand for health, in most cases, can not be measured directly. The empirical data available is utilisation or expenditure data, which is used as a proxy for actual demand. The utilisation data available represents demand for health care not the actual demand for health even though deductions can be made about actual demand

[22] Cullis and West (1979b)
[23] Fuchs (1972)

for health from such data. This theory is similar to Grossman where he implies that the demand for health care is derived from the demand for health, the difference here though is that Fuchs implies that individuals demand health for itself whilst Grossman implies that individuals demand for health is derived (increased healthy days mean increased leisure and work time).

In his work Fuchs analyses data from the USA, regarding health care utilisation, between 1947 and 1967. Using these data Fuchs is able to make economic inferences regarding the price and income elasticities of medical care. The paper claims that increases in price and income do affect the level of total health care demanded however, all possible increases (decreases) in prices or income should not be viewed in isolation, since factors such as population, prices and income or, GDP per capita will all affect the total demand for health care. Grossman's empirical work was a micro (individualised) analysis of the impact of factors upon health care demand (investment) whilst Fuchs work is more of a macro analysis. The implication (top-line) is that price of health care and income (or wage in Grossman's model) do affect demand for health care (derived from the demand for health).

2.3.7 How can Grossman's model be applied to the sight test

If we apply Grossman's model to the sight test as a health-improvement tool we can see that the commodity 'good eye sight' is produced from individual's own time and attendance at the optician for a sight test. Thus, the demand for a sight test is derived from the demand for 'good eye sight'.

Health in Grossman's model is both an investment and a consumption good; where health stock is a durable commodity and thus a component of the human capital framework, sight test is a health-improvement tool and by definition is durable. The individual will inherit an initial stock of vision, which after a point in time will begin to depreciate until the stock of vision is so low that sight loss is permanent, causing blindness. To maintain or increase the individual's level of vision the individual must invest in the sight test and its components.

If the shadow price of health increases as ages increases, then the depreciation rate of health will be increasing over the individual's life-cycle, as it does with vision – in that the decline in vision will increase as the individual ages at a rate whereby the sight test cannot reverse the vision loss but can go some way to aiding vision through glasses – then surely a free sight test for the elderly would make more economic sense, in terms of maximising utility from available infinite resources.

In terms, of the sight test, it would seem that the more educated an individual is so the more s/he would demand vision; increasing age leads to a decline in vision and a subsequent increase in expenditure in terms of sight tests spectacles and so forth. Since health has a positive wage elasticity, whilst health expenditure has a negative unearned income elasticity, one could extrapolate from Grossman's work that the higher an individuals wage so the greater his/her level of vision demand whereas an increase in income will

lead to a decline in vision. Therefore, as individuals retire, their wage will decline and so will their demand for sight testing.

Following on from Grossman's analysis of health care demand, it can be deduced that the quantity of sight demanded is negatively correlated with the shadow price of sight, which in turn is dependent upon many variables other than the price of a sight test. Grossman theorises that the shadow price will rise as age increases if the rate of depreciation of the stock of health, and in our analysis sight, increases over the life cycle as it will with individuals after a given time point, and will fall with education since the more educated individuals are more efficient health producers. It is possible that the shadow price of sight will also be related to factors such as wealth, wage rates, and other variables; the older an individual becomes, so the shadow price of sight will increase leading to a decline in the demand for sight, so providing a supportive statement for free sight testing to be provided for older individuals, or the over 60s.

Grossman poses a very articulate and in-depth analysis of the demand for health; wage, income and, prices are important factors when an individual is demanding improved vision. It is not possible to deduce the exact effect that a change in income or prices, among other factors, would have upon the individual's attendance for a sight test, however Grossman's model does imply that a reduction in the price of the sight test (especially for individuals on a lower wage) will increase their demand for the sight test, and so their vision stock.

2.3.8 Summary of the economic theory of health improvement

The individual is able to undertake health investment and consumption based upon his/her perceived knowledge of the current stock of health thus rationally make a utility maximising decision based upon his/her current state of knowledge. The level of sight-improvement that the individual will demand for his/her vision is dependent upon factors (income, market and non-market activities, depreciation) as well as price of the sight test.

2.4 Screening

2.4.1 Economic theory of screening

In this section we look at the economic theory of screening as a tool to detect health problems and apply this to the current issue – why is it necessary for the government to intervene in the sight test market buy providing partial provision.

In the following section we examine the Phelps model for the demand for screening very closely applicability to the sight test. Phelps model use the expected utility maximising consumer framework model and makes a major contribution to the understanding of the demand for preventative care and its relation to other goods that the individual will consume to maximise his/her utility; which is very relevant to demand for the sight test as a preventative tool.

2.4.2 Phelps Model

Phelps (1978)[24] looks at the expected utility maximising consumer, from which he develops a theory regarding the demand for preventative medical services. The value of preventative medical care is dependent on the value of the pure health gains – increase in utility, minimised loss of work time and, minimised expenses.

[24] Phelps (1978)

2.4.3 His assumptions

Phelps' model assumes that the consumer will rationally maximise his/her utility (utility is a function of health and other goods) subject to a fixed budget constraint (sum of wage and non-wage earnings), and an exogenous time endowment. Health is assumed to be a randomly distributed exogenous variable which is known by the consumer. The individual is able to increase his/her health level by undertaking preventative and acute medical care activities.

In his model, Phelps' assumes that the gain from preventative medical care can be quantified as medical expenses saved, that can be used in the consumption of other goods and so a useful indicator of welfare[25]. However, this measure could be misleading since market prices are likely to be used for medical care, which in the presence of medical insurance (of in the case of the UK the NHS), is an overstatement of the resources freed up for other use by the consumer; the cost saved from undertaking a preventative action will not be realised.

2.4.4 How does it all fit together

Phelps models the demand for acute and preventative medical care, where the individual maximises his/her expected utility subject to a budget constraint.

The budget constraint:

$$I = p_p P + p_m m + p_\chi \chi = N + w[T(H) - t_\chi \chi - t_m m - t_p P]$$

[25] Phelps (1978)

where;

P = preventative care; p_p = price of preventative care; t_p = time taken to consume P

m = acute medical care; p_m = price of m; t_m = time taken to consume m

χ = other goods; p_χ = price of other goods; t_χ = time taken to consume x

w = wage rate; N = non-wage-earnings

H = Health

T = Time

The individual's budget (I) is constrained by his/her total expenditure, split between expenditure on preventative care, acute medical care and other goods; the total expenditure is subject to the sum of wage and non-wage earnings that the individual has available. Non-wage earnings are an exogenous value whilst wage earnings are a factor of the 'free healthy' time available; 'healthy' time available is a factor of the exogenous value of total time available to the individual, multiplied by the individual's level of health; 'free healthy' time is simply 'healthy' time minus time taken by the individual to consume preventative care, acute medical care and other goods.

The individual's utility maximising function is based upon the likelihood of becoming ill, wage and non-wage earnings, and consumption of preventative and acute medical care as well as other goods.

$$E(U) = \pi\{U[\chi_1, H_1 + g(P)] + \lambda_1[-wT - N + (p\chi + w^*t_\chi)\chi_1 + (p_p+w^*t_p)P]\}+(1-\pi)\{U[\chi_2, H_2 - L + g(m,P)] +\lambda_2[-w(t-L(m))-N+(p_\chi+w^*t_\chi)\chi_2+(p_p+w^*t_p)P+(p_m+w^*t_m)m]\}$$

where;

P = preventative care; p_p = price of preventative care; t_p = time taken to consume P

m = acute medical care; p_m = price of m; t_m = time taken to consume m

χ_i = other goods; p_χ = price of other goods; t_χ = time taken to consume x

w = wage rate; N = non-wage-earnings; λ_i = marginal utility of income

H = Health; g(P,m) = health production function

π = probability of being healthy; $(1 - \pi)$ = probability of being sick

L = Time lost from sickness

T = Time

The amount of preventative medical care is selected only at the beginning of the period, and so is common to all states of the world, that means whether ill or healthy the individual is able to benefit from preventative medical care. The probability of an individual is a function of how much prevention the individual has.

The individual maximises his/her expected utility by maximising his/her probability of being healthy and minimising the probability of being ill. Maximising the probability of being healthy consists of maximising the individual's utility which is a function of being healthy, consuming

preventative care and, preventative health production function. Minimising the probability of being ill is a consists of maximising the individual's utility subject to health, and use of preventative and acute medical care, and a function of the marginal utility of income, which is based on the wage-time lost due to illness, non-wage earnings and the time taken to consume medical care (acute and preventative) and other goods.

Although Phelps highlights that there is widespread belief in the efficacy of screening, there is little evidence concerning the actual outcomes for screening patients[26]. After analysing the work carried out by Friedman et al, on the early diagnosis of breast cancer and the Forst screening trial comparing detection and morbidity rates between Naval and Army personnel, the data suggested that a common reduction in the price of both acute and preventative care would not necessarily increase the use of preventative medical care.

The data seemed to suggest that there is little or no effect of screening in medical use or health; one reason for this could be due to the sensitivity and specificity of the test procedures used, as well as a lack of confidence in the procedures held by doctors. For many diseases Phelps found that early detection may not necessarily change the course of the disease instead bringing out a false sense of improvement; since patient survival from the time of detection may appear to be increasing but instead, it is only an

[26] Phelps (1978) (p.200)

improvement in the detection. Also there may be a value of information benefit in screening that is not detectable by Phelps' model.

2.4.5 What are the implications for demand and price of health services

Essentially Phelps' model would seem to suggest that if there is an increase in preventative care then the individual will decrease the use of acute medical care, which means that if the price of preventative care decreases then there will be an increase in the demand for preventative medical care, given that the budget constraint remains constant, implying a reduction in acute medical care.

However the latter analysis carried out by Phelps suggests that this may not necessarily be the case, however this could be due to 'faulty' or incorrect measures of health benefit. However the analysis does imply that individuals may get an information benefit, from knowing that they are healthy.

2.4.6 How does the other literature diverge from Phelps model

A paper looking at the effect of a change in health cover for a certain health care intervention is that by Manning and Phelps (1979). Here they examine the prospect of an increase in the demand for dental care if this were to be placed on the NHI (National Health Insurance) bill under full or partial coverage, this is in a similar vein to the provision of the sight test under the NHS, free to the over 60's. The importance of looking at the price elasticity of demand for dental care, to see if a reduction in the cost will lead to an increase in demand and, subsequently, usage is considered along with the

importance of looking at services provided by a dental health check-up and availability of tools for appropriate dental health care promotion.

Manning and Phelps (1979) after reviewing, several models, found that demand is associated with income, price of treatment, and variables that will affect tastes for dental care, such as the presence or absence of dental disease. A review of the literature indicated, to Manning and Phelps, that income should be separated at source since this would allow for significant time costs associated with the consumption of dental care. For dental health care, the household production model indicates that demand may be different for one service than another. Individuals demand services of good teeth, including the ability to chew food, good appearance and the freedom of good teeth, these services can be achieved by preventive care, such as fillings/caps or extractions/replacements. Each procedure has a complimentary and substitution relationship in the production of dental health, thus it is important to assess the effect of various stimuli on demand for each service type.

In this paper Manning and Phelps considered the implications of providing dental care of the NIH program in late 1970's USA. Structurally, the paper goes through the literature already available on dental care demand, and then applies a model to 1970 general household survey data to find the relationship of price and income, amongst other variables to dental care demand. The findings of this paper are that the demand for dental care is highly price and income elastic, with full insurance resulting in a higher

demand for dental care and higher income also resulting in a higher demand for insured dental care. Manning and Phelps found that as the price of medical care tends towards 0, so the price elasticity of demand will tend toward zero. Thus, zero price will increase demand and reduce the sensitivity of demand to market price fluctuations.

Manning and Phelps paper suggests that as the price of medical care tends towards 0 so demand will increase; this seems to agree with Phelps that a decrease in medical care cost will increase demand, however Phelps model looks specifically at the preventative and acute medical care as separate but combined health factors, whereas Manning and Phelps examine dental care as a preventative medical care.

2.4.7 How does the model apply to the sight test

Thus in reference to Phelps's paper we can say about the sight test that it's value to the individual is dependent upon the value placed upon the improvement in vision, and inevitably the reduced hindrance in work and leisure activities. Therefore, using Phelps's model to interpret the sight tests, we could say that the gain for preventative medical care is, instead of medical expenses saved, annoyance and hindrance averted, work and leisure time saved. It is true, when considering a screening intervention that early detection of disease may give a false sense of confidence. This, however, may not be the case when screening for eye disease through the sight test.; since earlier detection will lead to treatment of the problem, be it glaucoma, cataract or a visual impairment, and so an improvement in the morbidity of the disease will be seen.

Phelps paper casts a shadow of doubt on the rhetoric that all health screening is necessarily a good thing. The review of studies in Phelps' paper highlights the potential pitfalls in allowing widespread screening to take place, greater allowance of medical screening will not necessarily mean that improved health and reduced medical expenditure will necessarily follow as given. However, it may also be that the methods by which the value of these preventative tests are assessed may be undervalued and there may be a need to improve the valuing techniques available so that all benefits incurred to the patients can be appropriately assessed. Phelps concludes that it is important to appropriately assess preventive medicine before employing, or promoting, its widespread use. This is assuredly relevant and important to consider when assessing the benefits and pitfalls of screening via the sight test.

Phelps's paper is an analysis of the overall preventive medical care market, which although not specific to the sight test can be used to indicate the effect of sight testing as preventive medicine. However, sight testing is essentially opportunistic screening whilst Phelps looks at all screening, not necessarily distinguishing between primary, secondary or opportunistic.

2.4.8 Summary of the economic theory of screening

Phelps (1978) and Manning & Phelps (1979) try to apply theory to preventive medical care. With Phelps looking at preventive medical care as a whole and Manning & Phelps examining more specifically the effect of placing dental care, either under full or partial coverage, onto the National Health Insurance program.

Both studies suggest that an increase in the preventative medical care price will lead to a decline in medical care demand.

2.5 Discussion

Analysis of the health improvement aspect of a two-pronged tool implies that if the marginal cost can be equated to the marginal benefit of health capital the individual will maximise his/her health stock, however reducing the price to zero will not necessarily imply an increase in demand as there are non-monetary costs that may deter the individual's decision. The analysis of the screening aspect of the two-pronged tool concludes that by reducing the price of screening there will be an increase in the demand for that screening; increased demand for screening leads to a greater likelihood of detecting diseases earlier thus causing a decrease in the use of acute medical services.

When analysing the sight test as a double-pronged tool, by improving vision and screening to capture possible vision threatening problems at an earlier stage, this can lead to an improvement in patient and society welfare. An improvement in vision will lead to an increase in the individual's utility (better vision means the individual is able to participate in economic and non-economic activities) and is able to maximise his/her production function.

There is no empirical data to support the case for sight testing. It is necessary to gather information regarding the factors that affect sight test attendance and this is carried out in the empirical section of this thesis.

There is no literature available that directly analyses the benefits of an opportunistic screening programme that also provides a direct health improvement, though literature on health care demand does provide a basis

allowing us to appreciate the factors that are relevant in the individual's decision matrix. The factors most relevant to the consumer are age, income, and price of health with which the individual will attempt to maximise utility.

Demand for a health care tool that provides a health improvement and a screening benefit is influenced by the price of the tool; however price equal to zero does not mean that demand will increase to infinity since there are other factors that will influence individual's demand.

2.6 References

Bachmann M, Nelson S. Impact of diabetic retinopathy screening on a British district population: case detection and blindness prevention in an evidence-based model. Journal of Epidemiology and Community Health 1998; 52: 45-52.

Barr N. The Economics of the Welfare State. London: Weidenfeld & Nicolson, 1987.

Cohen DR, Henderson JB. Health, Prevention, and Economics. Oxford: Oxford University Press, 1991.

College of Optometrists. Common Eye Problems. http://www.college-optometrists.org/public/disease.htm 1998.

College of Optometrists. What happens in an eye examination. http://www.college-optometrists.org/public/information/eye_exam.htm 1999.

Cullis J, Jones P. Public Finance and Public Choice: analytical perspectives. London: McGraw - Hill Book Publishing Company, 1992.

Cullis JG, West PA. The Economics of Health Care. Oxford: Martin Robertson and Company, 1979.

Daily Telegraph Oct. 12. Smart metals keep wearers in good shape. 1999; 41.

Department of Health. Number of Sight Tests and Vouchers paid for by HAs: by year. http://www.doh.gov.uk/public/pub_health_care.htm 2001

Department of Health. Statistical Bulletin: Ophthalmic Statistics for England -- 1985-86 to 1996-96. London: Government Statistical Service (GSS), 1996.

Eddy DM. Screening for Cancer. New Jersey: Prentice-Hall, 1980.

FODO. Optics at a Glance: 1996 Edition. : Federation of Ophthalmic and Dispensing Opticians (FODO), 1996.

FODO. Optics at a Glance: Year Ending 31 March 1997. : Federation of Ophthalmic and Dispensing Opticians, 1997.

Fuchs VR. The Growing Demand for Medical Care. In: Fuchs VR ed. Essays in the Economics of Health and Health Care. : NBER, 1972.

Goddard M. Sight Testing and the Role of Optometry. York: University of York, Centre for Health Economics: Health Economics Consortium, 1990.

Gottlieb LK, et al. Glaucoma Screening. A Cost-Effectiveness Analysis. Survey of Ophthalmology 1983; 28: 206-226.

Grossman M. The Demand for Health: a Theoretical and Empirical Investigation. : NBER, 1972.

Hawkins JM. Oxford Mini Dictionary. New York: Clarendon Press, Oxford University Press, 1992.

Holland WW, Stewart S. Screening in Health Care: Benefit or Bane? : Nuffield Provincial Hospitals Trust, 1990.

Kenkel DS. The demand for preventive medical care. Applied economics 1994; 26: 313-25.

Minassian, DC.; Reidy, A.; Coffey, M.; Minassian, A. Utility of predictive equations for estimating the prevalence and incidence of primary open angle glaucoma in the UK. British Journal of Ophthalmology 2000; 84(10): 1159-61.

Phelps C. Illness Prevention and Medical Insurance. Journal of Human Resources 1978; XIII (Supplement): 183-207.

Reidy A, et al. Prevalence of Serious Eye Disease and Visual Impairment in a North London Population: Population Based, Cross Sectional Study. BMJ, 30 May 1998; 316: 1643-6.

Reinstein D. Correctable undetected visual acuity deficit in patients aged 65 and over attending an accident and emergency department. BJO 1993; 77: 293-6.

RNIB. RNIB Campaigns - Better health, better Wales. http://www.rnib.org.uk/campaign/betwales.htm 2001.

Ryan M. Estimating the Effects of Prescription Charges on the Use of NHS Prescribed Drugs in England, 1979 -1985. Aberdeen: University of Aberdeen, 1989.

Sheldrick J, Sharp A. Glaucoma screening clinic in general practice: prevalence of occult disease, and resource implications. British Journal of General Practice 1994; 44: 561-5.

Sim F, Vafidis G. Seeing over the Money Barrier. London: The London School of Hygiene and Tropical Medicine, 1990.

Smeeth L, Iliffe S. Community screening for visual impairment in the elderly (Cochrane Review). The Cochrane Library, Issue 1 2001.

Smeeth L, Iliffe S. Effectiveness of screening older people for impaired vision in community setting: systematic review of evidence from randomised controlled trials. BMJ 1998; 316: 660-3.

Thornton J. Should health screening be private? London: Institute of Economic Affairs, Health, and Welfare Unit, 1999.

Tuck M, Crick R. Efficiency of referral for suspected glaucoma. BMJ 1991; 302: 998-1000.

Tuck M, Crick R. The cost-effectiveness of various modes of screening for primary open angle glaucoma. Ophthalmic Epidemiology 1997; 4: 3-17.

Wormald R, Rauf A. Glaucoma screening. Journal of Medical Screening 1995; 2: 109-14.

3 Impact of sight testing eligibility on uptake: Data Analysis

3.1 Introduction

A change in government policy in 1998 saw the introduction of sight testing to the over 60's population on the NHS. This paper assesses the potential impact of eligibility for a free sight test upon uptake and subsequent patient and society welfare.

Our initial intuition is that individuals' sight test uptake is directly correlated to their own eligibility status, where eligibility status is based upon health and economic factors. The sight test provides a level of information that will affect the future uptake of the sight test and this is analysed empirically, sight test eligibility inevitably impacts upon the uptake of the sight test it is important to assess to what degree eligibility impacts upon uptake. The relationship between the sight test and the risk of requiring glasses is important but, could not be empirically assessed within the current models. This paper examines the impact of eligibility for a free sight test and other socio-economic factors upon the uptake of sight tests.

The sight test is a useful two-pronged tool in vision-related problems. Not only is the sight test able to detect vision problems that are correctable via glasses (ie refractive error) but also is a useful tool in detecting the existence of eye diseases such as cataracts, glaucoma and so forth, that if left undetected and untreated, may lead to permanent blindness, but would

definitely lead to a reduction in welfare not only to the individual but to society as well.

Economic theory has examined the impact of prevention per se and, specifically in breast cancer (Kenkel) the demand for health care as a derived demand for health (Grossman) using the health capital framework, this has been further examined with paper one of this thesis. However, there has been no attempt to apply economic theory to the type of investment that is 'free-at-point-of-contact' but may create an additional cost/saving, this is an analysis that is very suited to the sight test when seen as a preventative tool. The most obviously connected area where work of this nature is within the dental arena. Literature related to the economics of dental care (Manning and Phelps 1979) relates more to the role of insurance to dental attendance, with full and partial coverage of dental care there is an increase in dental attendance.

We use the first seven waves of the British Household Panel Survey (BHPS). This general survey covers approximately 10,000 individuals in 5,000 households, per year and includes information on sight testing, as well as a range of other health and socio-economic factors. Within our analysis we examine the sight test consumption rate of individuals over the seven years within relation to specific health and socio-economic factors. We found that initial attendance for a sight test is significantly less likely (-0.53) if the individual is not eligible for a free sight test. However, eligibility has no

impact upon whether the individual will return for a repeat sight test two years after the initial attendance.

Therefore, we conducted an analysis, upon longitudinal data over seven years, examining the impact of eligibility, history and other socio-economic factors upon the uptake of sight testing where the risk of an additional cost is unknown. Section two presents the literature supporting the empirical specification for the model. Section three discusses the data, with the estimation results in section four. Discussion of the results in relation to the health prevention and sight testing market is reported in section five. The last section concludes with implications for public health policy.

3.2 *Modelling sight test uptake*

Researchers have typically been interested in the uptake of sight testing as part of a prescriptive screening programme put in place to pick up a certain type of disease, such as a glaucoma screening programme. However opportunistic screening as part of the sight test has not been examined. This literature has been analysed within the previous part of this thesis.

Screening, per se, or preventative care as part of the human capital framework model has been used to examine breast cancer screening, this is the framework applied within this analysis. This is not to say that the sight test and breast cancer screening will necessarily have the same outcomes avoided (ie eye disease avoided vs breast cancer and related mortality avoided). The framework underpinning Kenkel, with breast cancer screening, is the human capital framework, the same framework hypothesised to underpin the sight test as a health improvement tool, as discussed earlier within this thesis. Using Kenkel's model of breast and cervical cancer screening uptake we can try to assess those factors that may be relevant in the uptake of sight tests.

Kenkel's paper examines the demand for adult preventive medical care, with respect to early cervical and breast cancer detection, focussing on variables that are expected, after examining Grossman's work on the demand for health care, to be important determinants of demand for health capital.

Although Kenkel's paper looks at breast and cervical cancer screening within the US health care system, where insurance is required for both the preventive and curative aspects of health care, I propose to look at the UK health care system. Within the UK system the NHS may provide the sight test, depending on eligibility categories however, whether an individual is able to obtain a sight test (preventive eye care screening) upon the NHS does not necessarily determine an individual's access to curative care, since this element is always available upon the NHS at zero price for patients (aside from their opportunity cost – ie time, travel etc).

With the use of probit models Kenkel (1994) estimates the determinants of women's purchases of preventive medical services designed for the early detection of breast and cervical cancer. Estimates show that annual use seems to decrease with age, in other words that health and preventative medical care seems to decline as the pay-off period to investment shortens over time, it would seem therefore that these patients are being economically rational. The model also estimates that an increased level of education will lead to increased usage. In some cases it was found that the decision to visit the doctor is positively related to preventive care usage, however this effect in most cases was outweighed by the age and education effect upon the demand for preventive services. Another interesting finding is that curative care insurance does not dissuade preventive care use; in actual fact it seems that both types of care are complements to one another rather than substitutes. The data that Kenkel uses is data from mid-1970's and early 1980's USA; the variables used are age, education, type of insurance cover,

income, whether the individual is, or was, pregnant in the last year, health status, their physician's speciality, whether the physician is certified, and age of physician. The dependent variables were breast examinations within the year and pap tests within the year.

Kenkel's empirical work is supportive of the theoretical basis of the paper; the lifecycle effect will dominate the health risk effect. This is obvious when there is a decrease in the use of breast and pap tests as age increases, especially over 60, even though it is well-recognised that the risk of breast and cervical cancer actually increases as age increases. An increased level of education tends to give an increased use of preventive services; Kenkel finds that education to at least high school standard has a large effect on the uptake of preventive care, whilst additional education provides small marginal effects. The data does not support Kenkel's initial moral hazard theory that insurance for curative care will lead to a decline in the purchase of preventive care. There is also evidence to suggest that higher income will tend individuals to receive preventive care, however the individual's general health status is important in explaining curative care but, not necessarily preventive care demand. Also, it would seem that the physician's speciality might be an important determinant for use of preventive care. The relationship between the responsiveness of preventative care demand and the change in insurance coverage are examined, Kenkel hypothesises that there may be a resulting moral hazard problem that if individuals are insured for curative care and not for preventive care, then individuals may tend to purchase less preventive care.

The framework underpinning the model to be used in this analysis is derived from the human capital approach, which is the basis for both Kenkel's (1994) and Grossman's (1972) work. Within the human capital framework, the individual has an initial stock of knowledge or health (as is the case with the sight test) which s/he can increase in order to improve his/her productivity in the market and non-market sector of the economy. In order for the individual to fully realise their potential productivity gains, s/he must invest in education (or health). The cost of the investment includes the direct investment outlay on market goods and the opportunity cost of time required for investment. Within the health care framework the 'gap' between inputs of medical care and output of health is the household production function. The household production function is affected by the efficiency/productivity of the individual. Within the human capital – health framework – the individual can either consume their health capital, to provide utility (illness-free days or longevity), or can invest it, to determine the amount of time available for (non-) market activities. An individual will generally aim to maximise their utility by using a combination of consumption and investment activities.

Using the human capital framework as a basis, Grossman was able to apply the analysis to health capital and inputs. Within the analysis, human capital was seen to be depreciating over time, with the ability to carry out gross investment. Kenkel looks at Grossman's framework for analysis of demand

for health care and applies it in relevance to preventative and curative care demand, with respect to breast and cervical cancer.

Within the human capital framework, and in its application to health capital, age and education play an important role. The health stock that an individual has will begin to depreciate at an increasing rate after a certain age so; as an individual's age increases the return that s/he would have from investing in health capital reduces. Whilst, the depreciation rate on health stock will fall as education increases, since education leads to an increase in the efficiency of health production so that more health is obtained from a given amount of health inputs, so education increases the individual's return to investment. Age, due to the life-cycle effect, will lead to a decrease in investment whilst education, due to improved efficiency, will lead to an increase. Thus, with reference to Kenkel, we can develop a model for individuals' decisions to attend for a sight test.

Kenkel's model has also been used to econometrically analyse the screening choices in cervical cancer, where primary and secondary prevention is defined and screening is defined as secondary prevention. This analysis (Belkar 2004) also uses the human capital model and Kenkel's framework to examine the impact of awareness of cervical cancer with demand for screening.

The theory surrounding human capital places emphasis upon socio-economic factors as affecting individuals' decisions to consume or invest their capital.

Within the framework of sight testing, individuals are faced with a stock of inherited vision capital which they will consume over time, usually as age increases. The stock of vision capital can be added to by investment in the form of sight testing. The modelling carried out in this paper examines the individual's decision to attend for a sight test and invest in his/her vision capital.

3.2.1 Sight test attendance

The general theoretical framework used in this paper follows the format of the equation below, and looks at the social and economic factors affecting the individual's decision to attend for a sight test.

$$y = X\beta + \varepsilon$$

where:

y= net return to screening (dependent variable),

X = explanatory (independent) variables, social and economic factors,

β = coefficients,

ε = error term.

Within the model we examine the socio-economic factors behind the decision to go for a sight test, all independent variables are taken as at year one whilst, the dependent variable is taken for each year from years one through to seven. The model is estimated using maximum likelihood multinomial logit modelling also known as polytomous logistic regression.

3.2.1.1 Model:

$$y = \beta_0 + \beta_1 \text{ i.ajbstat} + \beta_2 \text{ afree} + \beta_3 \text{ afiyr} + \beta_4 \text{ aage} + \beta_5 \text{ arace} + \beta_6 \text{ asex} + \beta_7 \text{ i.aqfedhi} + \beta_8 \text{ asmoker} + \beta_9 \text{ amastat1} + \beta_{10} \text{ atenure} + \beta_{11} \text{ adriver} + \varepsilon_i$$

where;

$$y = X\beta + \varepsilon$$

y = 0,1,2,3,....7 - the number of years individual made a positive attendance in the specified time period

and X_i = aage = Age in first year of data-set;

arace = Ethnicity;

asex = Gender;

aqfedhi = Education level attained in first year of data-set;

asmoker = Smoker status in the first year of the dataset;

amastat1 = Marital status in first year of data-set;

atenure = Tenure status in first year of data-set;

adriver = Driver status in first year of data-set;

ajbstat = Job status in first year of data-set;

afree = Eligibity status in first year of data-set;

afiyr = Income in first year of data-set.

However, the 'y' variable is not directly observable thus, there is the need to use the discrete approach where the dependent variable is individuals' attendance for a sight test over a number of years (in this case seven), where the sight test attendance is separated into three categories – regular, occasional and, non-attendance over the given period. This separation of

attendance into occasional and regular is necessary to capture the effect that various socio-economic factors may play upon these attendance levels. We believe that similar to the role of socio-economic factors upon sight test non-attendance; these very same factors play an important role in the decision to attend on either an occasional or regular basis.

The following section reports upon the data used to analyse this model as well as the logic behind the selection of the relevant variables in relation to the human capital framework. This model, above, is not directly related the models reported previously (Grossman and Phelps), this model examines the impact of socio-economic factors upon the uptake or demand of the sight test – how do certain factors affect the uptake of a possible screening and vision-improving good, that is what is the net return to the (product of the) sight test (i.e. vision-improvement and preventative aspect) of certain socio-economic factors. Grossman's model examined the demand for health based upon the investment of health capital and depreciation of initial capital stock, whilst Phelps examined the demand for preventative medical care based upon the initial amount of preventative care, the time endowment, and wage and non-wage earnings.

3.3 Sight test, heath and socio-economic data in the BHPS

The data needed to estimate the model stated above require representative data from the UK population encompassing social characteristics and economic characteristics. The data used is obtained from the British Household Panel Survey (BHPS) longitudinal panel data, and has been merged and cleaned.

The data-set originally comprised of approximately 5,000 households, approximately 10,000 individuals, with the panel being asked the same questions in a longitudinal survey every year over a seven year period starting in September 1991. Individuals who left the household to start a new household, or new individuals joining households, after the survey began were kept in the data set.

However, after merging all seven years data, and deleting responses that were missing, don't know and proxy response, the data set now comprises 4, 700 individuals, with each individual responding each year. Of course, the use of only complete cases does introduce selection bias, where individuals leaving or joining the household, for whatever reason, have been excluded. This could be explored analysing the data (ie socio-economic) for these incomplete cases and comparing with the 'complete' data; such analyses are not conducted in this paper.

A number of different variables were created, the first being the variable STEST. STEST signifies whether the individual visited the optician regularly, occasionally or not at all. Each category is calculated using the individual's response in each year, this is then summed and segregated into regular attendance (if individual visited the optician three times or over in a seven year period), occasional (if the individual visited the optician once or twice in the seven year period) and never attended (which is self-explanatory). Across the entire sample 23% of individuals never attended for a sight test in the seven year period under consideration, 35% of individuals were occasional attendees, whilst 42% were regular sight test attendees.

Similarly, the variable AFREE depicts those individuals that would be eligible for a free sight test under the pre-1998 eligibility criteria, where eligibility was due to low income, unemployment status or health, where the only cost is opportunity cost of time.

We look at annual income, AFIYR, in order to assess the effect of income levels upon sight test attendance, here again the cost to the individual (aside from sight test fee) is the opportunity cost, where higher earning individuals may have a higher opportunity cost when attending. Conversely, the sight test is a normal good; hence an increase in income should increase sight test attendance.

ATENURE represents individuals property ownership/rental status which may play a role in drawing attention to the wealth effects that may contribute to the individual's decision.

The education level or AQFEDHI in the model, as Grossman and Kenkel suggest, will increase the individuals' demand for preventive care and hence the demand for sight testing; since education may increase the efficiency of health production, which would in turn reduce the shadow price of vision. Education also could increase the allocative efficiency of individuals given that they are better informed, and so can achieve higher returns to their investment, especially in terms of occupation. The effect that education may have upon the sight test attendance is also dependent upon the importance that sight test places upon other market and non-market activities.

AAGE is the variable representing age in the first year of the survey for all respondents, and was selected to be included in the model due to the impact that age is documented as having upon health capital in Grossman's work, health capital depreciates as age increases. This has been examined earlier in the theoretical section of this paper where age increases so the risk of developing a preventable visual problem increases thus there should be a tendency for increasing age to increase the individual's demand for sight testing. Though it may be that at a different age, an individual may have different health objectives and so may have different incentives for health capital investment, the lifecycle effect may mean that if individuals have a shorter pay-off period to their investment there may be a tendency for sight

testing demand to decrease with age. The relative importance and strengths of the health-risk and life-cycle effect would determine the effect that age would have upon sight test demand.

Ethnicity is depicted by ARACE, whilst ASEX represents gender. These social variables may be relevant in the analysis of sight test demand to indicate possible differences in importance placed upon health capital by different individuals. Individuals of Afro-Caribbean ethnicity are more prone to eye problems such as glaucoma (as discussed in the theoretical section of this paper) which is a sight-threatening disease, thus attendance for a sight test would help identify individuals more quickly, saving future discomfort for the individual and cost to the state (in terms of registered blindness if the problem has not identified soon enough).

The benefits of good sight are complementary to other activities, such as occupation and driver status, which are in turn complements to market and non-market activities.

If we are to claim that provision of the sight test upon the NHS, would improve the attendance for the test, then necessarily sight test attendance could be connected to job status, or AJBSTAT. Since job status would determine whether a) a person would have 'time' to attend for a sight test and, b) the importance that they place upon their vision (in terms of job status).

The variable ADRIVER depicts the individuals' driving ability, which would in turn affect individual's non-market, and in some cases, market activities.

Individual's smoking status is depicted by the variable ASMOKER; a non-smoker may be more inclined to be a regular sight test attendee since s/he is more interested in their visual health and investment.

AMASTAT1 is a social variables representing marital status which may affect the individuals' decision to attend for a sight test, and is in place for reasons of completeness.

3.4 Results

Throughout the analysis, the category "Never attended" is used to represent the comparison group. This, essentially, means that the coefficients produced, when the equations are run, will give an odds-ratio of relative probability between the comparison group and the other two possible outcomes. The coefficients produced can, also, lend themselves to the calculation of marginal effects, which will show the marginal effect of a unit change in the value of the regressor. Marginal effects will be calculated to show exactly that. The "z" values produced are similar in interpretation to the OLS "t" values, in that they show the significance of that specific variable to the explanation of the outcome.

It is also important, at this point, to highlight the statistical information that is presented in both models. At the end of Table 7we can see the χ^2 test statistic which shows the joint significance of the independent variables in each

model, except the constant. The test statistic is far greater than the critical values for the appropriate degrees of freedom at the 5% level of significance.

3.4.1 Sight test attendance

Table 7 below shows the mean effect, with standard errors, of given coefficients upon occasional and non-attendance for a sight test and regular or non-attendance, based upon the model below.

Analysis of the model reported significant results for the eligibility status, income, age, gender, education, driver and smoker status for regular sight test attendees as compared to non-attendees, whilst only age and gender were significant results for occasional versus non-attendees.

Table 7: M-Logit model of decision to attend for sight test

Dep. Var.: STEST= 1 non-attendee; STEST= 2 occasional; STEST= 3 regular				
	Occasional Attendance		Regular Attendance	
Job Status (Not Working)	0.1767	(0.1061)	0.0333	(0.1050)
Eligibility (eligible)	0.0749	(0.1508)	0.6320	(0.1444)*
Annual Income (£'000's)	0.0106	(0.0064)	0.0177	(0.0062)*
Age	0.0331	(0.0033)*	0.0553	(0.0033)*
Ethnicity – Black	-0.1121	(0.6830)	1.0364	(0.5776)#
Ethnicity – Asian	0.1107	(0.3229)	0.1730	(0.3260)
Ethnicity – Other	0.4448	(0.4191)	0.3066	(0.4246)
Gender (male)	-0.4320	(0.0909)*	-0.6896	(0.0910)*
Higher Qualifications	0.1851	(0.1738)	-0.0189	(0.1690)
A-levels	0.2017	(0.1929)	0.1811	(0.1874)
O-levels	-0.8077	(0.1699)	-0.2296	(0.1651)
Apprenticeship	-0.1603	(0.1910)	-0.4262	(0.1879)*
No Qualifications	-0.2168	(0.1757)	-0.6344	(0.1717)*
Smoker Status (smoker)	-0.0693	(0.0901)	-0.2000	(0.0910)*
Marital Status (married)	0.0464	(0.0951)	-0.1227	(0.0942)
Tenure Status (own property)	-0.0354	(0.1010)	-0.0237	(0.1014)
Driver Status (driver)	-0.1453	(0.1047)	0.2342	(0.1065)*
Constant	-0.7689	(0.2410)*	-1.4097	(0.2407)*
LR chi2(34)	520.25			
Prob > chi2	0.0000			
Log likelihood	-4696.4962			
Pseudo R2	0.0525			

STEST	Actual	Predicted
1	1074	1054
2	1640	1619
3	1986	1969
	4700	4642

* denotes 5% level of significance; # denotes 10% level of significance

Positive eligibility status leads to significantly increased likelihood of regular attendance than not at all, as would be expected; as the price of health care tends towards zero so demand increases.

Greater income is significantly related to an increased likelihood of regular attendance than non-attendance; this is related to the hypothesis that

individuals with a greater income face a greater opportunity cost with deteriorating sight thus have a greater return to regular attendance.

Males are significantly less likely to be occasional/regular attendees than non-attendees, this fits with the overall health literature that suggests females are more likely to consume health.

The model results show that age is a significant factor to occasional/regular sight test attendance; suggesting that as opposed to the life-cycle effect health risk may be a more relevant factor for individuals. Whilst lower educational levels (apprenticeship and no qualifications) have a significantly lower likelihood of attending for a regular sight which corresponds to the health capital framework – greater education means that individuals are more efficient health producers.

Drivers are significantly more likely to be regular sight test attendees than non-attendees; this would be relevant given that individuals that can drive are more able to attend the opticians. Also drivers may lose their jobs and become unemployed if they have poor sight and this would also be a motivating factor to attend for sight tests.

Individuals who are smokers are significantly less likely to be regular sight test attendees; suggesting that individuals who do not smoke may be more likely to be health efficient and so more likely to be sight test attendees, which was a hypothesised previously.

In the following section we evaluate the marginal effects of the above model, in order to do so the base case is given as below:

1. of white ethnicity;
2. of degree level education;
3. female;
4. single;
5. rent;
6. non-smoker;
7. non-driver;
8. employed/self-employed;
9. not eligible for a free sight test.

Age and income are both taken at their mean value, for this m-logit model.

3.4.2 Predicted probabilities for sight test uptake

Predicted probabilities or the effect of each socio-economic factor, whilst keeping all other factors constant, upon predicted probabilities of sight test attendance are examined with respect to the base case as stated above. Social factors are examined in Table 9 and economic factors in Table 8.

Table 8: Predicted probabilities of attendance for a sight test – Economic factors

Variables	Sight test attendance frequency		
	Never	Occasional	Regular
Eligibility			
Base case: No	0.257	0.438	0.304
Yes	0.206	0.364	0.430
Income levels			
£15,000	0.239	0.433	0.329
£20,000	0.227	0.433	0.340
£30,000	0.206	0.432	0.362
£40,000	0.186	0.430	0.384

An interesting factor is that which considers eligibility status. Individuals are eligible for a free sight test based upon income and health considerations. It is interesting to note the large effect that can be noted in the marginal effect of an individual's change in eligibility status from non-eligible to eligible, all other factors remaining constant. An eligible individual has a predicted probability of 0.430 of regular sight test attendance. The eligibility factor is not as pronounced in occasional sight testing though, non-eligible individuals are more likely to attend occasionally, and predictably it is this group that is most likely to not attend at all.

Income is not seen as being a significant factor in the occasional vs. non-attendees part of this model, whereas this factor is significant in the interaction of regular vs. non-attendees. The lower income gives a greater predicted probability of non-attendance, whereas a higher income gives a higher predicted probability of regular attendance. It is important to recognise that whether the individual is eligible for free sight tests or not,

income levels are important since the individual may not be eligible for assistance towards the purchase of glasses.

Table 9: Predicted probabilities of attendance for a sight test – Social factors

	Sight test attendance frequency		
Variables	**Never**	**Occasional**	**Regular**
Age			
18 yrs	0.469	0.383	0.150
25 yrs	0.399	0.412	0.189
35 yrs	0.307	0.441	0.251
45 yrs	0.227	0.453	0.321
55 yrs	0.161	0.446	0.393
60 yrs	0.133	0.437	0.429
65 yrs	0.111	0.425	0.465
75 yrs	0.073	0.393	0.535
Education			
Base case: Degree	0.154	0.368	0.478
Higher Qualifications	0.144	0.414	0.442
A-levels	0.132	0.383	0.485
O-levels	0.174	0.387	0.438
Apprenticeship	0.193	0.400	0.407
No Qualifications	0.211	0.419	0.370
Ethnicity			
Base case: White	0.229	0.349	0.422
Black	0.144	0.186	0.671
Asian	0.206	0.348	0.445
Other	0.173	0.404	0.423
Gender			
Base case: Female	0.174	0.418	0.407
Male	0.259	0.418	0.323
Smoker			
Base case: Non-smoker	0.267	0.439	0.294
Smoker	0.246	0.429	0.325
Driver status			
Base case: No	0.252	0.412	0.335
Yes	0.251	0.479	0.270

Age, which was significant in both the occasional attendees and regular attendees, we can see that those aged 18 and 25 yrs old are most likely to not attend at all whereas those aged 65 and 75 yrs old are the most likely to attend regularly. This however tends to go against the demand for health literature based upon the human capital framework (ie Grossman) and the

life-cycle effect where age is negatively correlated to investment, however the health-risk effect indicates that demand is positively correlated to age.

Health capital framework as applied by both Grossman and subsequently Kenkel suggest that increasing education would tend to increase health demand. Within our analysis education was only significant in Apprenticeship or No qualifications for regular attendance. In the above analysis educational level does not have a great impact on the marginal effects of attendance frequency, although those at a higher education level do tend to attend more regularly.

Gender is significant in both occasional and regular attendees versus non-attendance; males have a greater predicted probability of not attending for a sight test at all, whereas females have a greater probability of attending on a regular basis, all other factors remaining constant.

Smoker status is a significant indicator for likelihood of regular versus non-attendance. All other factors remaining constant we can see that the likelihood of either non-, occasional or even regular attendance is not vastly different between smokers and non-smokers, which does not correspond to the analysis previously suggesting other underlying factors.

The use of an M-logit model assumes independence of irrelevant alternatives. The inherent problem of the M-logit model is independence of irrelevance of alternatives (IIA), where the ratio of probability (in this model) between

never and regular attendance is irrelevant to whether occasional attendance is included or not. Table 10 below depicts the results of the IIA test for model one.

Table 10: Test of IIA results for model one

Test: Ho: difference in coefficients not systematic		
	(b-B)'[(V_b-V_B)^(-1)](b-B)	
Test one:	Is inclusion of occasional attendance relevant to ratio of probability between never and regular	
chi2(16)	13.68	
	Prob>chi2	0.6899
Test one:	Is inclusion of regular attendance relevant to ratio of probability between never and occasional	
chi2(16)	-4.12	
	Prob>chi2	<0
	Model estimated on these data fails to meet the asymptotic assumptions of the Hausman test	

where b = less efficient estimates obtained from mlogit; B = fully efficient estimates obtained previously from mlogit

The IIA test shows that for the model the null hypothesis, that the difference between coefficients is not systematic, stands since the χ^2 test is 13.68, which has a probability of 0.6899. Thus, implying the null hypothesis stands and the assumption of IIA is not rejected. If we examine whether the ratio of probability between never and occasional is irrelevant to the inclusion of regular attendance then the IIA test depicts a χ^2 value of –4.13, which has a probability of less than zero, suggesting that the difference in coefficients is systematic and the m-logit is not the most appropriate method to estimate the model, and the type of model used to analyse the data brings in bias. However the negative variance seen could be rectified by using a later version of STATA™ where a more robust version of the Hausman test is available.

Another model choice could be to use count data model; however the use of a count data model in this instance would mean the model is dominated by the number of people attending the optometrist on a regular basis due to underlying vision problems, such as glaucoma; thus examining the incidence of underlying medical problems (such as glaucoma) rather than the analysis of the impact of factors upon individual's decision to undertake a health investment, which is the focus of this paper and thesis.

3.5 The effect of sight testing history

The sight test also produces information, a relevant factor for possible repeat attendance. The following section examines the role sight test history will play upon an individual's decision to attend for a sight test in a subsequent year, by estimating a bivariate probit version of a sample selection model. The bivariate probit model is used since, as in section 3.4, the IIA test is rejected for the m-logit model; in theory both the m-logit and probit models are quite comparable.

We hypothesis that once an individual has been for an initial sight test they are better informed and so have a greater incentive to return. The individual's decision to attend for a sight test is dependent upon certain social and economic factors, as indicated in the previous section. Previous attendance could be a good indication of future sight test attendance, thus we look at initial 'first' attendance in relation to repeat attendance within two years. This set-up allows us to examine repeat attendance as an overall follow-up attendance (regardless of when individual attended), to include those attendees that are returning for 'follow-up' i.e. those that attend every two years, as recommended, as well as those that attend yearly – due to health/vision reasons (e.g. glaucoma or poor vision) rather than to include those that attend once at some point, for instance, year one and then don't attend again until some future point, for instance, year seven.

Unlike the M-logit model the bivariate probit does not assume Independence of Irrelevance of Alternatives (IIA), thus eliminating any inherent problems

that were recognised to possibly exist in the m-logit model. The initial sight test is the individual's initial investment in information; the individual is better informed post-test one. The impact of this information may play an important role in the individuals' decision to attend for further test.

3.5.1 Econometric model

The model run is based upon the maximum likelihood bivariate probit sample selection model, which is defined as the *heckprob* model in STATA™. For the selected sample, those with positive sight test attendance history, the individual's characteristics and eligibility plays an important part in the decision to attend for a sight test. The probability of initial attendance (ST) is a function of exogenous characteristics (X_1) as used in model one (except education) and education status (Y), as well as random term, (ε_1):

$$ST = f_2 (X_2, Y, \varepsilon_1) \qquad (1)$$

The probability of a repeat sight test attendance (RP) is assumed to be a function of the individual's exogenous characteristics (X_1), and a random error term (ε_1):

$$RP = f_1 (X_1, \varepsilon_1) \qquad (2)$$

The individual's characteristics as described by X_1 include the explanatory variables used in model one above, except education level, since the

individual's education level is important in getting them to make the initial investment in information after which it is the information that they have obtained that ill be a factor in their subsequent repeat attendances.

Since the dependent variables used in both the selection (1) equation and primary (2) equation are binary, a bivariate probit version of the sample selection model is estimated by maximum likelihood.

3.5.2 Results

Table 11 contains the results for equation one, the individual's decision to repeat attend within two years of their initial attendance. Table 12 reports the results for equation two the individual's decision to make an initial sight test attendance.

Table 11: Bivariate probit sample selection model $ST = f_2(X_2, Y, \varepsilon_1)$

Dep.Var. = Individual made an initial attendance for a sight test; Yes=1; No=0			
	Individual made an initial attendance for a sight test		
	Mean (SE)		Z-value
Job Status (Not Working)	-0.0307	(0.0597)	-0.51
Eligibility (eligible)	-0.4320	(0.0771)	-5.60*
Annual Income (£'000's)	–0.0000	(0.0000)	-1.57
Age	-0.0150	(0.0017)	-8.89*
Ethnicity – Black	-0.6304	(0.2518)	-2.50*
Ethnicity – Asian	-0.0744	(0.1924)	-0.39
Ethnicity – Other	0.2421	(0.2642)	0.92
Gender (male)	0.2369	(0.0516)	4.59*
Higher Qualifications	0.1233	(0.0805)	1.53
A-levels	0.0133	(0.0936)	0.14
O-levels	0.2410	(0.0824)	2.92*
Apprenticeship	0.2836	(0.0953)	2.98*
No Qualifications	0.3953	(0.0855)	4.62*
Smoker Status (smoker)	0.1439	(0.0544)	2.65*
Marital Status (married)	0.0821	(0.0519)	1.58
Tenure Status (own property)	0.0540	(0.0588)	0.92
Driver Status (driver)	-0.1218	(0.0601)	-2.03*
Constant	1.4753	(0.1328)	11.11*

*5% level of significance; # 10% level of significance

Table 12: Bivariate probit model – selected sample equation RP = $f_1(X_1, \varepsilon_1)$

Dep.Var. = Individual has attended the sight test within two years of first attendance; Yes=1; No=0			
	Individual has attended the sight test within two years of first attendance		
	Mean (SE)		Z-value
Job Status (Not Working)	0.0530	(0.0509)	1.04
Eligibility (eligible)	-0.0416	(0.0757)	-0.54
Annual Income (£'000's)	0.0000	(0.0000)	1.94#
Age	0.1260	(0.0021)	5.98*
Ethnicity – Black	0.0103	(0.2915)	0.04
Ethnicity – Asian	-0.0037	(0.1609)	-0.02
Ethnicity – Other	0.3008	(0.2020)	1.49
Gender (male)	-0.1814	(0.0486)	-3.73*
Smoker Status (smoker)	-0.0611	(0.0457)	-1.34
Marital Status (married)	0.0533	(0.0471)	1.13
Tenure Status (own property)	0.0101	(0.0491)	0.20
Driver Status (driver)	0.0075	(0.0508)	0.15
Constant	-0.2964	(0.1000)	-2.96*
Censored obs	4642		
Uncensored obs	712		
Wald chi2(12)	65.60		
Log likelihood	-4161.888		
Prob > chi2	0.0000		
Athrho	1.2430	(0.2896)	4.29*
Rho	0.8463	(0.8219)	
LR test of indep. eqns. (rho = 0):			
chi2(1)	12.59		
Prob > chi2	0.0004		

*5% level of significance; # 10% level of significance

Individual's eligibility status shows a significantly increased likelihood of initial sight test attendance; the individual is more likely to attend for a sight test (make the initial investment in information) if s/he is eligible for a free test. Eligibility is not a significant factor for repeat attendance.

Age is a significant factor in initial attendance; older individuals have a greater likelihood of attendance in year one. Older individuals are significantly more likely to re-attend within the two year period.

Gender is a significant factor in initial attendance, similar to the previous analysis, males are less likely to attend.

Education levels (apprenticeship and no qualifications) are significant factors in individuals attending for an initial sight test, interestingly the lower education levels (O-levels and lower) produce significant coefficients. This is not relevant in repeat attendance.

Individuals of Asian and Black ethnicity have a significantly lower likelihood of repeat attending whilst non-smokers are significantly more likely to be repeat attendees.

Initial attendance and repeat attendance within two years are positively correlated, as shown by the correlation coefficient ($\rho = 0.84$). This suggests that there is a positive 'random' error correlation, that 84% of the random error (white noise) can be explained by the relationship between the initial sight test attendance and repeat sight test attendance within two years.

The results from this analysis show that though initial attendance for a sight test is dependent upon a number of factors, specifically eligibility, repeat attendance is not dependent upon eligibility. Thus to get the individual

through the door (so to speak) a price equal to zero for the sight test is very relevant however to maintain the individual's repeat attendance is not dependent upon eligibility.

3.6 Conclusion

Sight test eligibility potentially impacts upon individual and society welfare by improving disease pick up rates.

Our initial intuition is that sight test eligibility is directly correlated to sight test uptake; eligibility is based upon health and economic factors.

Providing free sight tests to the over 60's is an equitable allocation of resources, with the potential to improve individual and society welfare; eligibility to free sight testing however does not seem to automatically suggest increased demand. Other factors, as yet unobserved within this analysis, are relevant. One such factor is the risk of having to make an additional payment (ie glasses) however data to analyse this in relation to attendance and sight testing is not available.

The government has an incentive to provide free sight testing, as this provides an efficient method by which to screen for sight threatening diseases and so avoid future potential costs. The adoption of a prescriptive mass screening programme is not a cost-effective option as has been reported in the previous section of this thesis.

The individual does not have an incentive to screen his/her sight for potential sight threatening disease, therefore the government provides free sight testing to increase efficiency and equity in the allocation of resources and reduce potential externalities in the form of a reduction in future costs to the NHS.

The literature surrounding the demand for sight testing and relevant factors is relatively scarce; literature on preventative health care demand assumes age, insurance, and education play important roles. The theoretical framework of the demand for sight test as a screening tool is based upon the paper developed by Kenkel (1994); using the human capital framework and probit modelling Kenkel (1994) estimates the determinants of women's purchases of preventative medical care services designed for the early detection of breast and cervical cancer. Breast cancer screening and vision screening are not identical, at least in the potential outcome avoided (cancer (leading to potential death) vs sight (leading to potential blindness)) however the latent, or asymptomatic nature of the disease is relevant for both to provide a potential efficient and equitable improvement for individuals and society.

If we are to assume that preventative care is a part of the health capital framework then, we should expect demand to be positively related to increasing education (since the more educated are more efficient producers of health), though the health-risk effect would suggest that increasing age lends itself to increased demand since the greater age the age so risk too will rise. However, if we take the life-cycle theory (as assumed by Grossman) then

increasing age would indeed lead to a decline in preventative care use, as the return to investment would decline.

The model found that increasing age tends to be positively correlated with the uptake of sight testing, whilst income is not relevant. Individuals with a lower educational status are less likely to invest in their sight. Eligibility also has a very significant impact upon the uptake of sight testing. However, eligibility does not have a significant impact upon repeat attendance two years post-initial attendance – suggesting that there are other factors, aside from price that are inherent in the individuals decision-making process.

The sight test will improve vision and allow the process of screening for eye diseases to occur. However, the potential impact of financial considerations upon uptake, such as the risk of glasses, can cause a decline in vision and a decline in disease identification; leading to a decline in welfare for the individual as well as society.

The analysis indicates that if we apply the sight test as a preventative tool then the human capital approach is a valid approach and eligibility for a free sight test, as well as increasing age, will lead to a greater likelihood of sight test uptake.

3.7 References

Belkar R., Fiebig DG., Haas M. and Viney R. Why worry about awareness in choice problems? Econometric analysis of screening and cervical cancer. Feb 2004

Grossman M. The Demand for Health: a Theoretical and Empirical Investigation. : NBER, 1972.

Kenkel DS. The demand for preventive medical care. Applied economics 1994; 26: 313-25.

Manning, William G. and Phelps, Charles E., "The demand for dental care," Bell Journal of Economics, 1979, 503-25.

Ryan M. Estimating the Effects of Prescription Charges on the Use of NHS Prescribed Drugs in England, 1979 -1985. Aberdeen: University of Aberdeen, 1989.

3.8 Appendices

This appendix looks at the descriptive statistics and the m-logit models that have been produced from BHPS longitudinal panel data, after the data set has been merged and cleaned.

The data-set originally contained approximately 5,000 households, with approximately 10,000 individuals, responding, every year, over a seven year period, starting in September 1991, with the panel being asked the same questions in a longitudinal survey every year. Individuals who left the household to start a new household or new individuals joining households, after the survey began, were kept in the data set.

However, after merging all seven years data, and deleting responses that were missing, don't know and proxy response, the data set now comprises 4, 700 individual responses, with each individual responding each year.

A number of different variables were created, the first being the variable STEST. STEST signifies whether the individual visited the optician regularly, occasionally or not at all. Each category is calculated using the individual's response in each year, this is then summed and segregated into regular attendance (if individual visited the optician three times or over in a seven year period), occasional (if the individual visited the optician once or

twice in the seven year period) and never attended (which is self-explanatory).

Variables	Original values	Created Values
Stest		never
		occasionally
		regularly
asex	male	male
	female	female
arace	white	White
	black-carribean	Black
	black-african	
	black-other	
	indian	Asian
	pakistani	
	bangladeshi	
	chinese	
	other	Other
ajbstat	self-employed	
	paid employed	Employed
	Unemployed	Not Working
	Retired	
	Family care	
	FT student	
	long-term sick/disabled	
	Maternity leave	
	gov't training scheme	
	Other	
aqfedhi	higher degree	Degree
	first degree	
	teaching qualification	Higher QF
	higher QF	
	Nursing QF	
	GCE A-levels	A-levels
	GCE O-levels	O-Levels
	Commercial QF, no O-levels	Apprenticeship
	CSE Grade 2-5	
	Apprenticeship	
	Other QF	
	No QF	No QF

asmoker	Smoker	Yes
	Non-Smoker	No
amastat	Married	Married/Co-habit
	Living as a Couple	
	Widowed	Not married
	Divorced	
	Separated	
	Never Married	
atenure	owned outright	Own
	owned with mortgage	
	local authority rent	Rent
	housing association rent	
	rent from employer	
	rent private unfurnished	
	rent private furnished	
	other rent	
adriver	yes	Yes
	no	No
afree	no	No
	yes	Yes

4 Impact of prescription charges upon uptake of GP consultations: Theoretical and empirical analysis

4.1 Introduction

Rising prescription charges, since their introduction in 1952, have led to suggestions that these are barriers to patient's prescription uptake; previous literature examining the impact of prescription charges and prescription utilisation and GP attendance have been undertaken and find a relationship to exist. This paper examines the relationship of prescription charges to GP attendance and prescription dispensation.

Our initial intuition is that prescription charges (and eligibility) will impact upon patients' attendance at the GP for a consultation, whilst also impacting upon the patient's decision to dispense prescriptions, since a possible future cost may well deter the individual making the initial health investment (GP consultation). Current published literature reports an effect (which can be differentiated as an effect between exempt and non-exempt individuals) upon prescription utilisation brought on by charges, either their introduction or increase. We hypothesise that prescription charge increases will not lead to a reduction in GP attendance for either eligible or ineligible patients, prescription charges will not impact upon the health screening impact of the GP consultation process, but implying that either a reduction in GP prescription writing or a concerted decision by ineligible individuals not to dispense a prescription will occur. Previous literature reported a decline in GP attendance when prescription charges were introduced however literature does not report an impact upon GP attendance with prescription

increases, it is our hypothesis that individual's GP attendance will not be affected by prescription charge increases as this is the screening element of the health capital investment, whilst prescription dispensation may be affected by individual's opting for an over-the-counter cheaper medication. It is important to examine and accept the joint or contingent relationship that exists between GP attendance and prescription dispensation. In order to do so it is necessary for us to examine the type of people that will be deterred from consulting the GP and what will deter these people from consulting their GP – such as the prescription charge.

Therefore we conduct an analysis upon three year longitudinal data, examining the relevant factors that influence GP attendance – these being such factors as prescription charges, free prescription eligibility, free prescription ineligibility, and socio-economic factors -- after which we examine what factors affect the individuals' decision to dispense their prescription. Section two reports the existing literature supporting the empirical model structure. Section three presents the background to present day prescription charging. Section four presents the empirical model, with section five reporting the data. Section six reports upon the estimation results. Discussion of the results, along with implications for current policy and resource allocation are carried out in section seven.

4.2 Review of relevant existing literature

Relevant published literature reports upon one of three hypotheses, reported by Ryan (1989), that could occur due to the introduction/increase of prescription charges either (1) individuals are not going to their GP (prescription charges are a barrier to attending the GP) or, (2) GPs are not prescribing as often as they should (prescription charges are a barrier to prescription writing) or, (3) individuals are not dispensing their prescriptions (prescription charges are a deterrent to individuals completing their health purchase). However, it is not possible to separate the literature into these three distinct hypotheses. Therefore the following literature review is presented in a dichotomous manner – examining the literature that reports the impact of prescription charges upon GP attendance and then reporting the impact of prescription charges upon prescription utilisation (be this prescription writing or prescription dispensation).

The first of the hypotheses stated above is that the prescription charge (whether introduced or increased) will be a barrier, or deterrent, to individuals consulting their GP (GP attendance). Below we discuss the extent to which this hypothesis is explored within current published literature.

A study carried out by Hardman (1965), over a two month period in 1965, found that the abolishment of prescription charges in 1965 led to an increase in GP consultations. The study was carried out in a GP practice population of 3,229 between January 4th and February 28th 1965; the specific time period was chosen to examine the impact of abolishing the prescription charge on February 1st 1965.

There was a 9.7%[27] increase in GP consultations after the abolition of prescription charges. The study is not, however, of a sufficient length to take account of seasonal factors such as the prevalence and impact of serious colds/illnesses, it is possible that the winter of 1965 was an exceptionally bad one for picking up illnesses and this could be a factor in the increased GP consultations experienced post prescription charge abolition; the study was not able to assess the long-term impact of prescription charge abolition. The increased frequency of prescriptions post prescription charge abolition suggests that the increase in GP consultations may not have been a wasted resource – ie increased prescription frequency suggesting a legitimate illness/problem, however this is a short term assessment of the impact of prescription charge abolition and a longer term assessment is important.

Leck (1972) conducted an analysis on 30,000 GP practice patients in early 1968 and then again in 1969 (the actual months analysed are not reported) – prescription charges were re-introduced in June 1968. The analysis was carried out by comparing GP attendance over two months, one early in 1968 and one a year later. Leck examined the impact of re-introducing prescription charges on GP contact and prescription dispensation. Re-introducing charges did not seem to reduce contact. Leck's analysis suggests that within the two months studied prescription charges did not change the number of GP contacts, but this relationship is only examined briefly over a two month period which is not of a sufficient length to fully assess the longer term impact of prescription charge introduction.

[27] Hardman (1965) p.356

Therefore, we can summarise that current published literature is inconclusive upon the impact of prescription charges on GP attendance; Hardman reports an increase in GP attendance occurring with the abolishment of the prescription charge, whilst Leck suggests that the re-introduction of the prescription charge would have no impact upon GP attendance. These two studies are small and of limited length; the results must be taken with caution when assessing the impact of prescription charges upon GP attendance.

Below we examine the impact of the prescription charge (whether introduced or increased) upon prescription utilisation; whether the GP will be deterred from writing prescriptions (especially for those individuals that are not eligible for a free prescription) and whether patients will be deterred from dispensing their prescription (becoming non-compliant) is discussed with relevance to current published literature, below.

Leck (1972) as mentioned before conducted an analysis on 30,000 GP practice patients in 1968 and 1969 over a two month period, the analysis found that re-introduction of charges in 1968 decreased prescription dispensation by one-tenth. Prescriptions dispensed to some of the patients during the second survey month were compared with those dispensed in the first survey month; the proportion of prescriptions un-dispensed was relatively high among non-exempt patients, especially for those living in relatively poor areas or who were seriously ill. Re-introducing charges did not seem to reduce GP attendance but Leck (1972) suggests that it may be a combination of reduction in prescriptions dispensed and

reduction in frequency of prescribing that may have contributed to the decline in prescriptions when charges were re-introduced, therefore prescription charges had an impact on prescription utilisation. The study, however does not examine the relationship over a longer period of time in order to fully appreciate other relevant factors, such as seasonality, its possible that the two months examined were exceptionally good months in terms of illnesses prevalent in the community.

Another study examining the impact of price upon prescription utilisation is Lavers (1989), using monthly data from January 1971 to December 1982 found that demand for prescription utilisation was price responsive during the specified period. Lavers reports upon an analysis of the effect of prescription charge increases between January 1971 and December 1982 (during which period prescription charges changed four times) upon prescription utilisation (the volume of non-exempt prescriptions actually dispensed – without a segregation for prescription writing or dispensation), the cost of their ingredients and the level of reported morbidity. Lavers solely looks at the effect of prescription charge increases on those patients who are not eligible for free prescriptions, i.e. the non-exempt adult population.

Lavers employs three equations to assess the size of the effect of price changes upon the demand for and the cost of supplying prescriptions; the dependent variables are the number of non-exempt prescriptions actually dispensed (Q), the average cost (to the NHS) of a prescription (C) and, morbidity (or the number of employed individuals that are certified sick and off work) in the population (M);

the independent variables are price of prescription (P), the relative price of substitutes (R), measure of income for the population (Y), the level of sickness benefit (B), a change in eligibility for prescription charges and, a seasonal variable (S). Demand (utilisation – writing or dispensation) is dependent upon morbidity and the exogenous variables whilst, morbidity and cost only depend upon the independent variables. Lavers found that the evidence implies that the demand (utilisation) for prescriptions was responsive to price in the specified period, with an estimated elasticity of between -0.15 and -0.20. However, there was little evidence to support the idea that price rises lead either to the prescription of significantly higher quantities of drugs or the prescription of relatively more expensive items.

Another study examining the impact of prescription charges upon prescription utilisation, though over a shorter time period (1979 – 1985) is Ryan (1989) finding that prescription charge increases during this period (a 900% nominal increase or a 490% real increase) led to a long-term decrease in utilisation of NHS prescribed drugs in the non-elderly non-exempt adult population. However, as expected the increase in prescription rates did not affect the non-elderly exempt adult population (ie those that do not have to pay for their prescriptions).

Ryan (1989), using data between 1979 and 1985, set out to examine three hypotheses that prescription charge increases will lead to a) patients being deterred from consulting their doctor; b) patients consulting their doctor but the number of prescriptions written being reduced; and c) patients consulting their

doctor and receiving the same number of prescriptions but failing to get them dispensed (increase in non-compliance). Of these hypotheses, Ryan was only able to examine the second hypothesis due to a lack of data availability. At the time of Ryan's work data on changes in GP consultation rates over time was not available and the data available regarding patient non-compliance did not address prescription charges as a possible explanation for this non-compliance. Given that, as Ryan claims, a substantial proportion of the population are exempt from prescription charges[28] then following economic theory an increase in prescription charges shouldn't affect utilisation within the exempt group, thus Ryan's analysis focuses upon the adult non-elderly non-exempt and exempt populations separately. Ryan models the relationship between prescription charge increases and use of NHS prescribed drugs, using exempt and non-exempt utilisation, dependent variables, and the prescription charge, price of substitutes, exempt and non-exempt income, retail price index, number of GPs, morbidity proxy, limited list and monthly dummy variables and a trend variable as independent variables.

Ryan concludes that the increases in NHS prescription charges, over and above the inflation rate, between the years 1979-1985, did affect the utilisation of NHS prescribed drugs for the adult non-elderly non-exempt population, where a 10% increase in the prescription charge would give rise to a short-run 1.06% and long-run 2% decrease in utilisation. Ryan supposes that although these effects are small, they are not simply one-off decreases but will continue through time; hence prescription charge increases over time may not be minimal decreases in

[28] Ryan (1989) p.11

utilisation. The increase in prescription charge did not effect the utilisation of NHS prescribed medicines in the adult non-elderly exempt population.

However, problems inherent with the relationship between the prescription charge and NHS prescribed drugs utilisation (as examined by Ryan) include the concept of consumer sovereignty (the individual's decision not to dispense a prescription rather than the GP's decision not to write a prescription), asymmetry of information (the GP will have more knowledge of the individual's health level and possible consequences of delaying treatment) and increased quantity per prescription (GPs may increase the number of tablets prescribed so the individual does not have to re-present prescriptions so reducing the overall cost to the individual). This type of relationship does not allow for consumer sovereignty so the individual patient is not allowed to judge what is best for him/herself (i.e. whether it is best to be prescribed the drug or not - that decision lies in the hands of the GP). Within health care there is, also, an asymmetry of information between the GP and patient, where the patient does not have all the relevant information upon which to base an informed decision and so is dependent upon the GP; the observed consumption of NHS drugs is dependent not only upon patient demand characteristics but factors affecting GP behaviour, too. In dealing with this problem, Ryan (1989) modelled the study as one which examines the utilisation of NHS prescribed drugs and not patient demand for NHS drugs. A further problem is that GPs may increase the number of drugs prescribed per prescription, if they feel that patients will be deterred from obtaining prescriptions, however Ryan's study looks only at the number of

prescriptions dispensed rather than the quantity of drugs per prescription, and this is made explicit.

Ryan's work points out that there is a reduction in the utilisation of NHS prescribed drugs. It is not possible to say whether this reduction is of unwarranted or frivolous prescriptions, it may be the case that prescription charges lead to a delay in individuals receiving the medication required. Ryan's paper concludes that an increase in prescription charges will decrease utilisation and increase revenue; however, there will be those that will delay receiving medication. It can not be said that the decrease in utilisation will necessarily lead to a decline in an individual's health, since further research is required into this area.

O'Brien (1989) examines the relationship between the change in prescription charges and prescription drug utilisation in England between 1969 and 1986, and finds that drug utilisation has negative charge elasticity for the patients that pay for their prescriptions and positive charge elasticity for patients that do not pay for their prescription at the time of drug dispensation.

O'Brien states the same three hypotheses as Ryan (1989) that could occur with a prescription charge increase and tries to examine these using monthly time-series data for England between 1969 and 1986 taken from the General Household and National Morbidity Surveys; however, O'Brien felt that these data were not only very ad hoc but also point estimates of consultations and not consistent time-series data to test the effect of prescription charges on GP contact. Data on the

number of prescriptions written, aside from prescriptions dispensed, was not nationally available. Given the data issues mentioned above, O'Brien felt it was difficult to sufficiently test both hypotheses (b) and (c) and assess whether increases in charges affect GP behaviour or patient compliance. O'Brien felt that the marginal increase in prescription charges would not be sufficient to significantly reduce consultation rates since GP consultations typically had a high time-price component attached. Although increased charges could affect GP prescribing behaviour especially where over-the-counter (OTC) medicine was available as an alternative.

O'Brien sets up a model to examine the relationship between charges and utilisation. O'Brien sets up two models, chargeable prescriptions and exempt prescriptions at time t, both affected by financial variables (X_i) such as real prescription charges, real price of substitutes and real income; demographic variables (Y_j) such as the number of elderly in the population, the number of under 16s, those of working age and a morbidity indicator of new claims for sickness and invalidity benefit; seasonal dummies (S_k) to allow of monthly variations and shift dummies (D_l) to allow for the extension of exemption categories in 1974, the introduction of no-charges contraceptives in 1975 and the introduction of limited prescribing in 1985.

$$\text{Chargeable prescriptions}_t = \beta_0^{C} + \Sigma\beta_i^{C}X_{it} + \Sigma\beta_j^{C}Y_{jt} + \Sigma\beta_k^{C}S_{kt} + \Sigma\beta_l^{C}D_{lt} + U_t$$

$$\text{Exempt prescriptions}_t = \beta_0^{E} + \Sigma\beta_i^{E}X_{it} + \Sigma\beta_j^{E}Y_{jt} + \Sigma\beta_k^{E}S_{kt} + \Sigma\beta_l^{E}D_{lt} + U_t$$

where U_t is the error terms and the intercept coefficients β_0^C and β_0^E are to be estimated for both equations; both dependent variables are measuring prescribing volume, rather than per capita utilisation. Though per capita utilisation would be desirable O'Brien highlights the data problems that prevented this being carried out, such as free prescription eligibility not simply being a case of demographics but also related to disease groups and contraceptive use.

The results of O'Brien's model give price elasticities of between -0.23 and -0.64 for the non-exempt equations for the overall period; whilst the price elasticity for the exempt equation is +0.17 for the entire period between 1969 and 1986. These elasticities tallied with O'Brien's hypothesis that as the per-item prescription charge increases so prescription utilisation for non-exempt, non-pre-payment population decreases, and non-chargeable utilisation increases.

However, one of the main problems with the model is that there is no way to disaggregate those individuals who are eligible for a free prescription from those with pre-payment certificates (those individuals that do not pay for their prescriptions at dispensation but, would have purchased a pre-paid 'season ticket' for their prescriptions), the prescription data used reports data for those dispensed that are paid for and those that are not paid. Also, the inability of the dependent variables to estimate per capita prescription instead of volume of prescription is a problem of the model.

A study carried out by Schafheutle (2002) found that most patients are cost-conscious when it comes to managing their condition with medicines. The study

was conducted via six focus groups and 31 patients all taking prescriptions (for dyspepsia, hay fever or hypertension, or hormone replacement therapy) and recruited through three community pharmacies in the North-west of England. For those people that have to pay for their own prescriptions but are on a low or moderate income, cost is an important factor. Some of the methods of dealing with the cost implications were either by not getting items dispensed, using a lower dosage so the supply lasts longer or delaying dispensing the prescription.

To summarise the current published literature examining the impact of prescription charges on prescription utilisation (prescription writing as well as prescription dispensation) we find that prescription charge increases will lead to a decline in prescription utilisation, for the adult (non-elderly) non-exempt population. Current published literature does not provide a conclusive answer as to whether utilisation declines due to the decision that the GP may not have written a prescription – advising the individual to obtain an over the counter (OTC) drug or even increasing the amount of medication per prescription or whether the increase in prescription charges cause an increase patient non-compliance. The literature reviewed above does suggest an impact on utilisation of prescriptions with Schafheutle's qualitative analysis reporting that cost is an important factor, when non-exempt individuals, are given a prescription, whilst Leck suggests that prescription charges may affect patients' level of non-compliance but does nor offer conclusive evidence to that effect.

All the papers that have been addressed in this review so far examine the effect of prescription charges upon certain elements of the three hypotheses stated

above, though not offering a conclusive answer to any on their own or linked. Prescription charges may or may not affect GP attendance; prescription charges are likely to impact, in a predictable manner, the prescription utilisation; however the impact of prescription charges upon compliance is inconclusive.

The hypothesis of this paper is to examine the impact of prescription charges upon the patient's decision to employ the GP as his/her agent (patient's demand for GP attendance), and upon the GP's decision (as the agent) to provide a prescription and the final decision, by the patient (as the principal) to be compliant or not.

As with the previous literature in this area it is not necessarily clear how to dissect the GP's decision to provide the patient with a prescription or not from the individual's decision (once s/he has the prescription) to having a prescription dispensed or not (compliance).

4.3 Background to prescriptions

Before launching into a model to try to examine the impact of certain factors upon GP attendance and the characteristics that may be relevant to prescription dispensation and writing, we have reported the background to prescription charges below.

Prescription charges were introduced on June 1st 1952. The charge began at 1 shilling per prescription form, increasing to 1 shilling per item on 1st December 1956. On March 1st 1961 the prescription charge was increased to 2 shillings per prescription item, however in 1965, (February 1st) the prescription charge was dropped, but re-introduced on June 10th 1968 at a charge of 2 shilling and 6 pence.

Since 1968 the prescription charge has been increased, to 20 pence (1st April 1971), and 45 pence (16th July 1979), the prescription charge was increased twice in 1980 to 70 pence (1st April) and then £1 (1st December). Since 1982 the prescription charge has been increased yearly (1st April) by 20 pence (1982-84), by 40 pence (1984 – 1990); by 25 pence (1990) and 35 pence (1991 – 1992). In 1992 the charge was £3.75, yearly increases at 50 pence each year continued until 1995 when the charge was £5.25. In 1996 the charge was increased by 25 pence, followed by an increase of 15 pence in 1997 and 1998 when the charge was £5.80. From 1999 to date the charge increased at a rate of 10 pence each year to the current charge of £6.60 (2005), these charges are all nominal increases, in real terms the charges have increased from 1979 to 1985 by 490%

(Ryan (1989)) and by 355% between 1980 to 1998 (Hitiris (2000)). Figure 1 shows the prescription charge history.

Certain individuals are eligible for a free prescription, this eligibility is based upon individuals being over the age of 60 or under the age of 16 (or 16-18 and still in full time education), war pensioners (who require prescription for their pensionable disability), pregnant women and women who have given birth in the last 12 months, persons with specified medical conditions, individuals and their partners receiving Family Credit, Working Families Tax Credit, Income Support, Income based Jobseekers Allowance, Disability Working Allowance, Disabled Person's Tax Credit, or those who qualify on the basis of low income. Contraceptives are also free from prescription charge.

Figure 1: History of prescription charges from 1971-2003

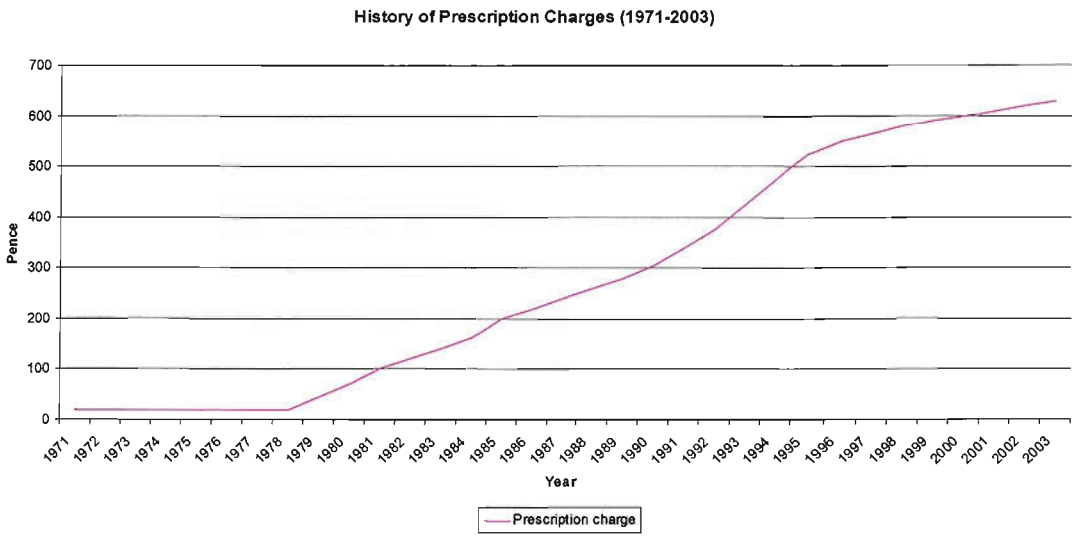


Figure 2 and Figure 3 display the history of GP consultations from 1972 to 2000, for males, females, and all individuals (aged between 0 and over 75) and for 16

to 64 year olds specifically. We can see that the GP consultations are relatively steady for individuals aged between 16 and 64 with a drop in 1998.

Figure 2: History of total average GP consultations from 1972 to 2000

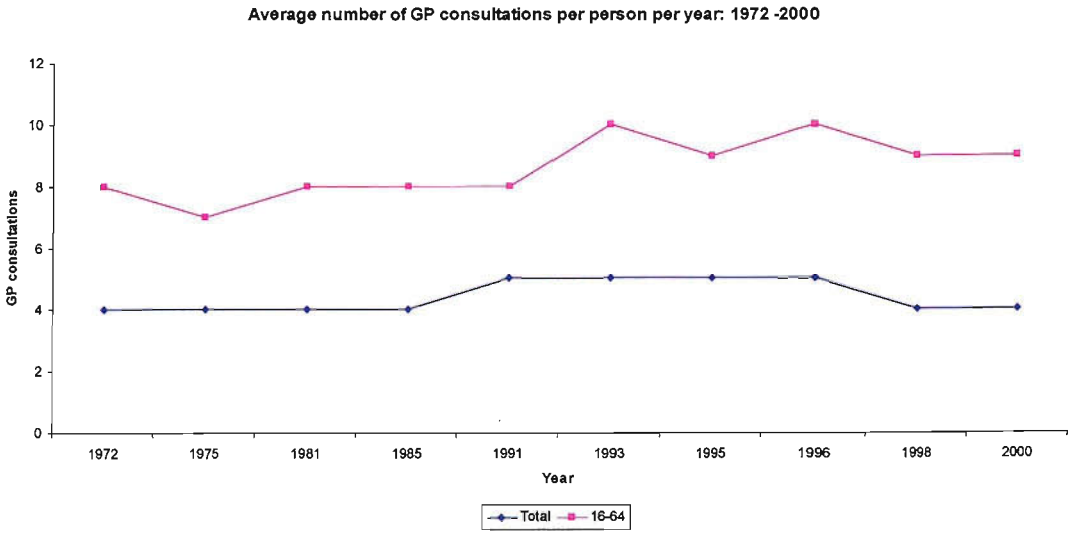
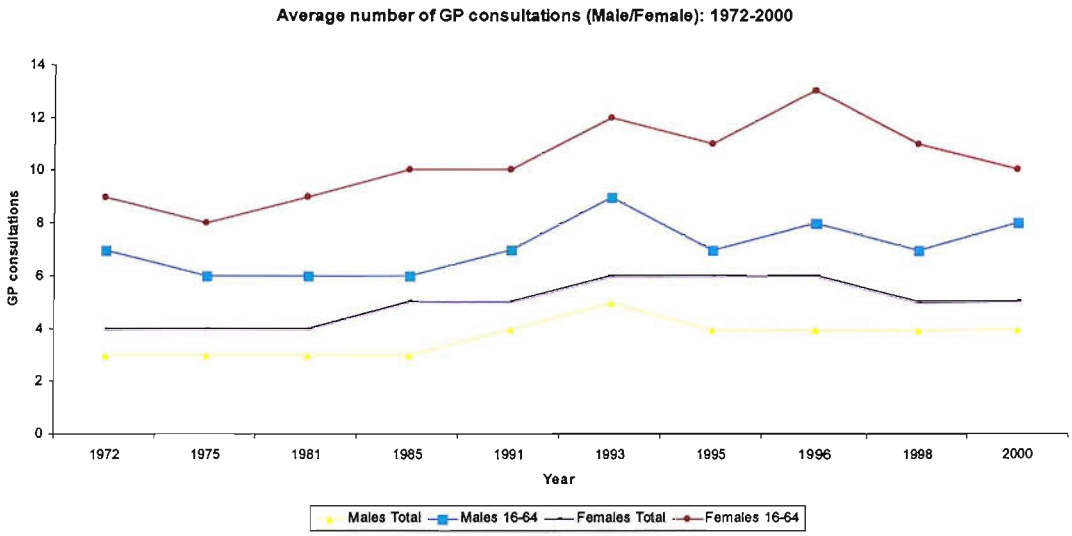
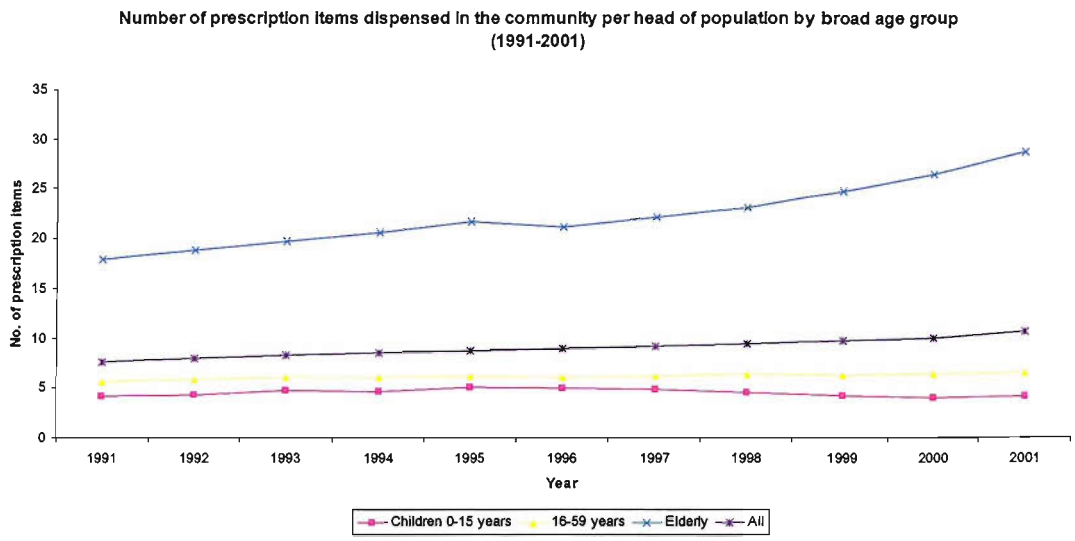


Figure 3: History of males/females average GP consultations from 1972 to 2000



The number of GP consultations for males (second curve from the top – blue) and females (top curve – brown) aged 16-64 increased from 1975 to 1996, with a drop in 1995, followed by a decline from 1996 onwards. The number of GP consultations for all males (bottom curve – yellow) and all females (second curve from the bottom - purple) increased from 1981 steadily, with a slight drop in 1995 for males and 1996 for females, after which the consultations remained the same.

Figure 4: Number of prescriptions items dispensed in the community per head of population b broad age group, 1991-2001



The number of prescription dispensed per head of the elderly population (top curve – blue) steadily increased from 1991 onwards, whilst the number of prescriptions for children (lowest curve – pink), 16-59 year olds (second curve from the bottom – yellow) and all individuals (second curve from the top – purple) increased slightly but remained steady overall.

The data depicted in Figure 2 show us that as prescription charging (Figure 1) increased over the years, the overall GP consultation attendance has remained relatively constant however consultation rates for all 16-64 year olds depicted an increasing trend, indicating that prescription charge increases did not have a major impact upon attendance. In Figure 3 both male and female individuals (both the whole group and the 16-64 year old age group) show an upward trend between 1972 and 2000. Figure 4 represents the prescription dispensation rates over the period 1991 to 2001, the children (0-15 year age group) and the adult (16-59 year age group) show a relatively constant trend throughout the period,

whereas the elderly age group (over 60 years old) shows a steep upward trend. A top-line level analysis of the data shows that as prescription charging increases so the average GP consultations per person have experienced an upward trend, dispensed prescriptions (overall) experienced a slight upward trend with the elderly experiencing a steep upward trend in prescription dispensation. However, this analysis does not take into account any potential impact that eligibility for free prescriptions may have had upon the average GP attendance rate.

Current evidence based on the current published literature and top-line data suggests that the increase in prescription charges does not impact attendance for a GP consultation; the abolishment of the prescription charge though did impact the GP consultation as expected whilst the re-introduction of the prescription charge does not impact the GP attendance. Whilst the prescription charge has been increasing over the years the data shows that the impact upon prescription utilisation has been relatively minimal. The literature indicates that the rise in prescription charges will lead to a decline in the utilisation of prescriptions especially for the non-exempt population.

4.4 Theoretical approach and Empirical model

As we have suggested above there are three factors that could be affecting prescriptions either; 1) individuals are not going to their GP or; 2) GPs are not prescribing as often as they should; 3) individuals are not having their prescriptions dispensed.

The papers reviewed in the existing literature section all report an effect (and some differentiated this effect between exempt and non-exempt individuals) upon prescriptions brought on by charges, either prescription charge introduction or increase. Ryan (1989) and O'Brien (1989) reported a difference in effect, of charges, upon exempt and non-exempt individuals, with a decrease in the prescription rate of the non-exempt but no change (or even increase) in the exempt population. Therefore, one could assume that given that there is no effect of the prescription rate in the exempt population, GPs are not changing their prescribing habits but, patients are either refraining from GP contact or they are simply not dispensing prescriptions written.

We know from the literature available that prescription utilisation has declined, due to prescription charge introduction and increases, however it is important to establish which part of the process this decline lies within. Data linking prescription dispensation to prescription writing is not readily available, thus the aim is to examine the initial part of this process, the effect of prescription charges on GP consultations by the non-exempt population. This has been briefly looked at by Hardman (found consultations in the specific practice looked at increased

with the abolition of charges) and Leck (found GP consultations did not change when charges re-introduced in the two month survey period).

The GP is employed by the patient as his/her agent, in order to provide information regarding current health state and potentially provide further interventions to tackle any health problems that may be occurring. However, the GP is also an agent for the government in that s/he uses the consultation to screen the patient for possible asymptomatic illnesses (illnesses the patient does not know about) and thus reduce the potential impact upon society.

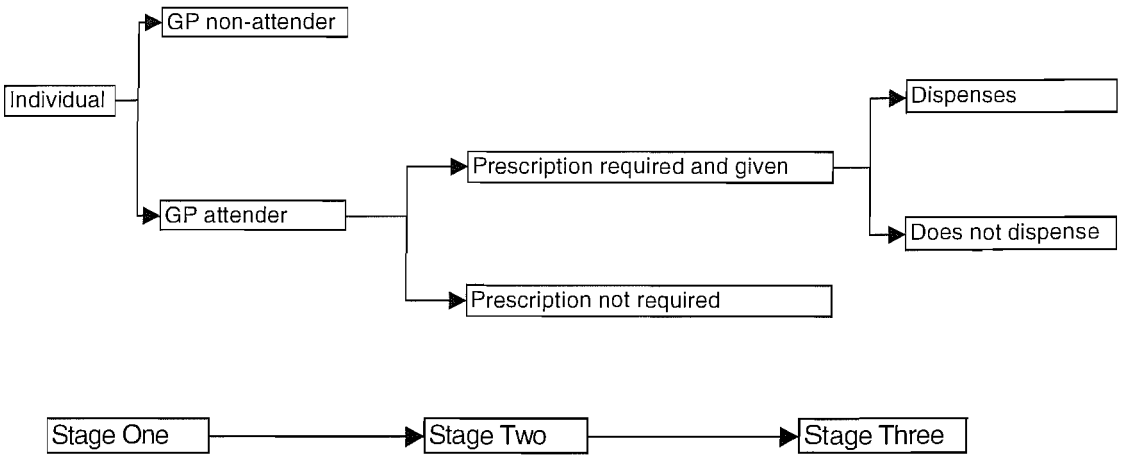
Primarily from society's perspective the GP is an agent to screen and reduce any potential impact of an (as yet) asymptomatic disease. GP attendance is free at the point of contact; however there is the potential to receive a prescription from that attendance that may incur a cost, a potential barrier. Policy has been developed that allows certain types of individuals to obtain free prescriptions, potentially reducing the barrier to GP attendance.

In order to maximise the potential welfare gain to society from GP attendances, it is necessary to establish those individuals that are likely to attend the GP and ascertain the type of individuals that are likely to not attend due the prescription charge.

Therefore we can say that the relationship between the (initial) GP consultation and (final) prescription dispensation is one of a contingent commodity. The individual's attendance for a GP consultation is contingent upon the individual's

perception of receiving a prescription, especially if we assume that all individuals going to the GP (who require a prescription) will be written one, since once the prescription is received a decision to use the prescription or not needs to be made.

This can be depicted diagrammatically as:



We have four states of the world, 1) Do not go to the GP; 2) Go to the GP and do not receive a prescription 3) Go to GP receive and dispense the prescription and 4) Go to the GP, receive but do not dispense prescription. At each stage of the process, as depicted above, the individual may gain information and have a cost. Stage one (individual goes to the GP) the individual only has the cost of time or travel, but gains information on his/her illness at stage two (prescription written or not) the individual will either be written a prescription or not, at the third stage (dispense prescription or not) the non-exempt individual has a (prescription) cost. Data reporting GP attendance and the prescription charge is available, which will inform the possible relationship between prescription charges and GP attendance

visualised in stage one, specific data relating prescription charges and dispensation is available for stage three and this has been examined previously as reported in the literature section previously, for the second stage the link between prescription charges and prescription writing is not readily visible and any analytical relationship would be tenuous and indirect. Therefore, this paper will examine the impact of prescription charges upon GP attendance directly as well as examining the link between attending the GP and dispensing a prescription.

This diagrammatic representation does, of course, ignore another additional stage which is whether the patient then actually takes the medication as intended by the GP or not, 'is the patient concordant or non-concordant?' The decision to concord with medication directions or not may not necessarily relate to whether the individual pays for his/her medication, unless the medication requires a number of courses of medication and non-concordance with the instructions means the patient extends the length of medication. However such an analysis, within the remits of the data available was not possible and was not conducted.

Basic economic theory would suggest that if a good has no price then the quantity produced in the market is dependent simply upon supply. In the case of GP consultations the quantity of GP consultations available in the market is determined by the supply of consultations at the given time, which may be limited by seasonality, or busyness of the week. However the amount of good demanded within the market when price is zero is not infinite but rationed by means of time; if the good is a normal good as the individual's income increases so demand increases, however as income increases so demand may decrease due

to the opportunity cost of time spent queuing for GP consultation. Therefore, I believe that the demand of GP consultations is affected by factors such as symptoms as well as risk of prescription charge, and opportunity cost of time.

Using Ryan's model of prescription charges and their effects, as a backdrop, we assess the effect of prescription charges upon individuals' attendance to a GP. However, this is further extended to assess the effect of prescription charges on GP prescription writing and the effect upon prescription dispensation. This section will try to model the economic model (for eligible and ineligible patients) in three parts firstly looking at the effect on GP attendance, on prescription writing and on dispensation of socio-economic factors and prescription charging.

The framework underpinning the models to be used in this analysis is derived from the human capital approach, the same as that used to underpin the sight testing analysis and, is based upon Grossman's (1972) work. Within the human capital framework, the individual has an initial stock of knowledge or health (as is the case with the sight test) which s/he can increase in order to improve his/her productivity in the market and non-market sector of the economy. In order for the individual to fully realise his/her potential productivity gains, s/he must invest in education (or health). The cost of investment includes the direct investment outlay on market goods and the opportunity cost of the time required. Within the health care framework the 'gap' between inputs of medical care and output of health is the health production function. The health production function is affected by the efficiency/productivity of the individual. Within the human capital - health framework, the individual can either consume their health

capital, to provide utility (illness-free days or longevity), or can invest it, to determine the amount of time available for (non-) market activities. An individual will generally aim to maximise their utility by using a combination of consumption and investment activities.

Using the human capital framework as a basis, Grossman (1972) was able to apply the analysis to health capital and inputs. Within the analysis, human capital was seen to be depreciating over time, with the ability to carry out gross investment depreciating also. This paper examines the hypothesis stated previously in terms of Grossman's framework for analysis of demand for health care and applies it in relevance to first and second stage linked-care demand, in other words where the first stage is essentially free, with the second stage subject to a fee, depending upon the individuals' circumstances, or contingent commodities.

Within the human capital framework, and in its application to health capital, age and education play an important role. The health stock that an individual has will begin to depreciate, at an increasing rate, after a certain age, so as an individual's age increases the return that s/he would have from investing in health capital reduces. Whilst, the depreciation rate on health stock will fall as education increases, since education leads to an increase in the efficiency of health production so that more health is obtained from a given amount of health inputs, so education increases the individual's return to investment. Age, due to the life-cycle effect, will lead to a decrease in investment whilst education, due to improved efficiency, will lead to an increase.

The theory surrounding human capital places emphasis upon socio-economic factors affecting individuals' decisions to consume or invest their capital. Within the framework of GP consultations, individuals are faced with a stock of inherited 'general health' capital that they will consume over time, usually as age increases. The stock of health capital can be added to by investment in the form of attending for a GP consultation (making an investment), and purchasing a subsequent prescription. Thus, for the individual there are two costs attached to adding to/maintaining their health capital, the cost of visiting the GP (initial 'required' investment) and the cost of purchasing the prescription (second 'possible' investment). The modelling carried out in this paper examines the individual's decision to attend for a GP consultation and invest in his/her health capital. Health capital investment can be estimated by the demand for health care, so we can say that the individual's decision to attend for a GP consultation is a proxy or the individual's demand for health and so health capital investment. The general theoretical framework used in this paper follows the format of the equation below, and looks at the social and economic factors affecting the individual's decision to go for a GP consultation.

$$y = X\beta + \varepsilon$$

where:

y_t = net return to GP consultation (dependent variable),

X_t = explanatory (independent) variables, social and economic factors,

β_t = coefficients,

ε_t = error term; t = time, static in the first model and dynamic in the second.

However, the y variable is not directly observable thus, there is the need to use the discrete approach where the dependent variable is individuals' attendance for a GP consultation over a number of years (in this case three years, with two weeks in each year), where the GP attendance is separated into attendance and non-attendance over the given period.

The empirical section first examines stage one as described in the theoretical section above, how the individual's GP attendance is affected by socio-economic factors (qualification, ethnicity, marital status, age, gender, car status, tenure status, and perceived social support).

Stage two of the theoretical section, does the GP write a prescription to the individual, assumes that if the patient is written a prescription then s/he will use the prescription; this model examines the impact of socio-economic factors on whether or not the individual is on a prescription (given that they have been to the GP – this is to remove the impact of repeat prescriptions). Repeat medication could bias the analysis, by over representing the prescriptions given and aligning these to the number of GP visits, thus causing an over-estimation of the relationship between GP visits and prescriptions given.

Analysis of stage two is a model based upon the maximum likelihood bivariate probit sample selection model, previously identified as the *heckprob* specification in STATA™. The probability of a prescription (P) is assumed to be a function of the individual's exogenous characteristics (X_1), and a random error term (ε_1):

$$P = f_1 (X_1, \varepsilon_1) \qquad (1)$$

The individual's characteristics as described by X_1 include the explanatory variables used in models in stage one above.

For the selected sample, those with GP attendance, the individual's characteristics, and their eligibility play an important part in the individual's prescription status. The probability of GP attendance (GP) is a function of exogenous characteristics (X_1) as used in model two and tenure status (Y), as well as random term, (ε_1):

$$GP = f_2 (X_2, Y, \varepsilon_1) \qquad (2)$$

Since the dependent variables used in both the selection (2) equation and primary (1) equation are binary, a bivariate probit version of the sample selection model is estimated by maximum likelihood.

Stage three of the theoretical section, does the patient use the medication, is modelled in a similar way to stage two, using similar socio-economic variables (the dependent variable is whether the individual is currently on prescription medication).

Analysis of stage three of the model is based upon the *heckprob* specification model in STATA™. The probability of being on medication (M) is assumed to

be a function of the individual's exogenous characteristics (X_1), and a random error term (ε_1):

$$M = f_1 (X_1, \varepsilon_1) \qquad (3)$$

The individual's characteristics as described by X_1 include the explanatory variables used in model six above.

For the selected sample, those with prescription written in the last two weeks, the individual's characteristics, and their eligibility play an important part in the individual's prescription status. The probability of prescription writing given that the individual attended the GP for a consultation (P) is a function of exogenous characteristics (X_1) as used in model one and tenure status (Y), as well as random term, (ε_1):

$$P = f_2 (X_2, Y, \varepsilon_1) \qquad (4)$$

4.5 Data

The data needed to estimate the models stated above require representative data from the UK population encompassing social characteristics and economic characteristics. The data used is from the cross-sectional 'Health Survey for England' data-set which is not longitudinal panel data but, data collected over a number of years asking the same questions at the same point in time to a similar mix of individuals; however the individuals are not the same year after year. Our models use the data for years 1998-2001, assuming that the data is connected between patients in 1998, 1999, 2000 and 2001, and estimate a time-series model. It is questionable whether the pooling of such data across the years is appropriate, given that there is the potential that individuals' attitudes may alter over time and thus lead to different behaviours coming through the data. In this case the data used was collected over a four-year period and it was felt that, though attitudes towards have changed, the change would not have been as sudden as to alter individual behaviour so dramatically as to cause significant issues with data pooling.

After merging and deleting responses that were missing, don't know and proxy response, the data set comprised approximately 25,000 individual responses. However, after removing the over 60 and under 18 population, which are those individuals that receive free prescriptions for reasons other than health or income there were ~22,000 non-elderly adult (18 – 60 years old) individuals left for analysis in 1998-2001 data set.

A number of different variables were created, the first being the variable 'GP'. 'GP' signifies whether the individual visited the GP (in the last two weeks) or not. Both across the entire sample and in each year (1998 to 2001) 84% of individuals attended the GP.

'Prescription' is a variable that reports whether the individual has been given a prescription in the last two weeks after visiting the GP, whilst 'Medication' indicates whether the individual is on prescription medication, excluding contraceptives. 'Charge' is indicative of the prescription charge each year, between 1998 and 2001, applicable to each patient based upon eligibility status.

'GP_Prescription' is a variable that reports upon those individuals that attended for a GP consultation in the previous two weeks and received a prescription in the previous two weeks. This variable was created to deal with the small number of patients that would receive a prescription in the previous two weeks but did not attend the GP for a consultation in the previous two weeks, ie those individuals receiving repeat medication. Of course

'Eligible' variable is created by separating eligible and ineligible patients; patients are identified as eligible for a free prescription, where eligibility is due to low income, unemployment status or, health (individuals diagnosed with cancer, diabetes, epilepsy, stroke, or heart attack). Individuals eligible for a free prescription will have a 0 charge applied.

'Heligible' variable is created by identifying patients as eligible for a free prescription, where eligibility is due to health (individuals diagnosed with cancer, diabetes, epilepsy, stroke, or heart attack).

'Seligible' variable is created by identifying patients as eligible for a free prescription, where eligibility is due to low income or unemployment status.

Annual income ('000s), 'income', is examined in order to assess the effect of income levels upon the dependent variables, here again the cost to the individual (aside from prescription fee) is the opportunity cost, where higher earning individuals may have a higher opportunity cost when attending. Conversely, if GP consultation is a normal good an increase in income should increase GP attendance.

'Age' is the variable representing age of patients in each year in the survey within the cross-sectional model, in the time-series model the age in 1998 is used. This variable was selected to be included in the model due to the impact that age is documented as having upon health capital in Grossman's work, health capital depreciates as age increases. Thus when age increases so the risk of developing a health problem increases, there should be a tendency for age to increase the individual's demand for GP attendance. Also, at a different age, an individual may have different health objectives and so may have different incentives for health capital investment, the lifecycle effect may mean that if individuals have a shorter pay-off period to their investment there may be a tendency for GP attendance to decrease with age. The relative importance and

strengths of the health-risk and life-cycle effect would determine the effect that age would have upon GP consultation demand.

Ethnicity is depicted by 'ethnic', whilst 'sex' represents gender. These social variables may be relevant in this analysis to indicate possible differences in importance placed upon health capital.

The education level or 'QF' in the models above, Grossman predicts, will increase the individuals' demand for health care and hence the demand for GP consultations. Since, education may increase the efficiency of health production, which would in turn reduce the shadow price of health. Education is also a proxy for occupation or earning level and so the greater the individual's education level so the greater the opportunity cost of being ill and so seeking preventative care; a greater education level also implies individuals are more efficient information processors, since they are better informed, and achieve higher returns to their investment. The effect that education may have upon the GP attendance is also dependent upon the importance that general health places upon other market and non-market activities.

If we are to assess the impact of prescription charges upon GP consultations and examine whether prescription charges do affect attendance for a consultation, then necessarily GP attendance could be connected to job status, or 'Employment'. Since job status would determine whether a) a person would have 'time' to attend for a consultation and, b) the importance that they place upon their general health (in terms of job status).

The benefits of good health are complementary to other activities, such as education, occupation, and leisure activities, which are in turn complements to market and non-market activities.

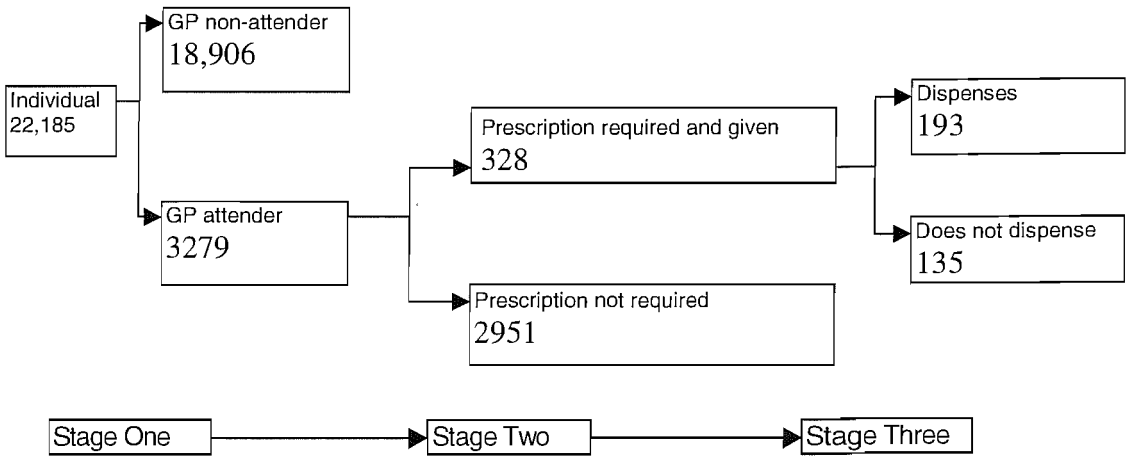
Variables 'marital' and 'tenure' are social variables representing marital status and individual's property ownership/rental status; these may affect the individual's decision to attend for a consultation, and are in place for reasons of completeness. However, the 'tenure' variable may also play a role in drawing attention to the wealth effects that may contribute to the individual's decision. 'Socsupport' is a variable that measures the individual's perception of the level of social support that they receive.

4.6 Results

This section looks at the results produced by the three-stage models above, using the M-logit model specification, and STATA™. The Health Survey data set was used for analysis. The data set originally contains over 25,000 different individuals, responding in years 1998 to 2001. However this data set is not longitudinal so individuals, though being asked the same questions, are not the same year after year.

After merging three years data, and deleting responses that were missing, don't know and proxy response, removing the over 60 and under 18 population, who receive free prescriptions for reasons other than health or income, ~22,000 individuals left for the 1998-2001 data set. The diagram below shows the proportion of patients that are applied to each stage (stage one to three).

Figure 5: Diagrammatical representation of the number of individuals eligible for analysis at each stage



Within each model, the dependent variable is described with the discrete values being 0 (Negative – i.e. either not attended, not given a prescription medication or, not on prescription medication) and 1 (Positive).

Throughout the analysis, the category "Negative" is used to represent the comparison group. This, essentially, means that the coefficients produced, when the equations are run, will give an odds-ratio of relative probability between the comparison group and the other possible outcome. The coefficients produced can, also, lend themselves to the calculation of marginal effects, which will show the marginal effect of a unit change in the value of the regressor. Marginal effects will be calculated to show exactly that. The "z" values produced are similar in interpretation to the OLS "t" values, in that they show the significance of that specific variable to the explanation of the outcome.

It is also important, at this point, to highlight the statistical information that is presented in both models. At the end of each table we can see the $\chi 2$ test statistic which shows the joint significance of the independent variables in each model, except the constant. In both cases the test statistic is far greater than the critical values for the appropriate degrees of freedom at the 5% and 10% levels of significance.

The analysis of these models is not conducted separately for exempt and non-exempt patients, reasons for this essentially centre around the dilution of data – separating the data into these to distinct populations caused a dilution of the effect of factors upon the dependent variable, therefore the decision was based

upon a 'power' argument; separation of the patients would cause the power of the analysis to reduce.

Stage One

This model in effect is trying to assess 'stage one' of the GP consultation process previously highlighted; stage one, in this dynamic relationship, examines the determinants of GP attendance, using socio-economic variables and prescription charges. The model is estimated using probit modelling.

$$GP = \beta_0 + \beta_1 \text{ Eligibility*Charge} + \beta_2 \text{ Heligible} + \beta_3 \text{ Seligible} + \beta_4 \text{ Income} + \beta_5 \text{ QF} + \beta_6 \text{ Employment} + \beta_7 \text{ Ethnic} + \beta_8 \text{ Marital} + \beta_9 \text{ Age} + \beta_{10} \text{ Sex} + \beta_{11} \text{ Tenure} + \beta_{12} \text{ Socsupport} + \varepsilon_i$$

where;

GP = GP consulted, by the individual, for themselves in the two weeks previous to interview;

Charge = Prescription charge for each year, 1998-2001, applicable to each patient dependent upon eligibility status;

Heligible = Eligibility dependent upon health;

Seligible = Eligibility dependent upon income and employment status;

Income = Annual net income ('000s);

QF = Highest qualification achieved – Degree, NVQ3/A-level, NVQ2/O-level/NVQ1/CSE/Foreign/Other, None;

Employment = Economic activity status – Employed, Unemployed, Economically Inactive (incl.retired);

Ethnic = Ethnicity – White, Black, Asian, Other;

Marital = Marital status – Single, Married/Co-habiting, Separated/Widowed/Divorced;

Age = Age;

Sex = Gender – Female, Male;

Tenure = Individual's tenure status – Own property, Part rent and part mortgage/rent;

Socsupport = Individual's perceived social support status – No lack, some lack, severe lack;

The reference individual (ie base-case) used in this (and subsequent models) is:

1. White;
2. Female;
3. Employed;
4. Degree qualified;
5. Owns own property;
6. Ineligible for free prescriptions (both social and health eligibilities), and
7. Perceives no lack of social support

We can see that charge, eligibility based upon income and unemployment status, income employment status (ie not employed), and gender (male) are all significant factors in this model.

Table 13: Probit model of individual's previous two week GP consultation status.

Dep.Var. = Individual has visited the GP for a consultation within the last two weeks; Yes=1; No=0			
	Mean (SE)		Z-value
Charge	-0.0650	(0.0232)	-2.80*
Health Eligibility	-0.1455	(0.1405)	-1.04
Social Eligibility	0.2748	(0.1191)	2.31*
Income	-0.0016	(0.0008)	-2.03*
QF_2	0.0038	(0.0339)	0.11
QF_3	0.0186	(0.0416)	0.45
QF_4	-0.0933	(0.0717)	-1.30
Employment_2	-0.1707	(0.0734)	-2.32*
Employment_3	0.2232	(0.0370)	6.04*
Ethnic_2	-0.0544	(0.0605)	-0.90
Ethnic_3	-0.0343	(0.0441)	-0.78
Ethnic_4	-0.0806	(0.1169)	-0.69
Marital_2	-0.0404	(0.0392)	-1.03
Marital_3	0.0634	(0.0513)	1.24
Age	0.0025	(0.0016)	1.57
Sex_1	-0.2560	(0.0296)	-8.65*
Tenure_2	0.0372	(0.0362)	1.03
Socsupport_2	0.0069	(0.0329)	0.21
Socsupport_3	0.0582	(0.0411)	1.41
Constant	-0.7411	(0.1527)	-4.85
LR chi2(19)	420.93		
Prob > chi2	0.0000		
Log likelihood	-5170.7674		
Pseudo R2	0.0391		

*5% level of significance; # 10% level of significance

Table 13 indicates that the charge has a significant effect on GP contact with individuals facing a charge being significantly less likely to attend for a GP prescription. Eligibility based upon income and unemployment status is a significant factor with eligible individuals significantly more likely to attend for a GP consultation than ineligible individuals, whilst eligibility based upon health

reasons is not a significant factor for individual's attending for a GP consultation. Males are significantly (26%) less likely to attend the GP for a consultation than female. Employment status is a significant factor in GP attendance, with unemployed individuals being 17% less likely to attend the GP than employed individuals and economically inactive individuals are 22% more likely to attend for a GP consultation than if s/he is employed, where inactive individuals also include pregnant ladies as well as individuals on disability benefits (social eligibility).

Therefore based on this we could conclude that unemployed males are less likely to attend for a GP consultation than older economically inactive females; with eligibility (based on income and unemployment status) being a very significant factor for increased attendance.

Stage Two

The econometric models in this stage of the theoretical model examine the relationship of prescription status with GP attendance taking into account socio-economic factors. This model attempts to grasp 'stage two' of the GP consultation and prescription utilisation relationship previously explained. In this section we look at the 'will the individual be given a prescription if s/he visits the GP'. Of course, at this stage this relationship requires us to make an assumption that if the GP writes a prescription then the patient will use the prescription (this third stage of use or not will be examined in the next section).

Table 14 contains the results for the sample selection model, where only those individuals who have attended the GP for a consultation in the last two weeks are examined. The results of the sample selection equation (ie those individuals with the dependent variable being whether they have been to the GP for a personal consultation in the last two weeks or not) have not been reported below as these are identical to those reported in Table 13.

In Table 14 reports that there are no significant factors in determining whether the individual has been given a prescription within the last two weeks given that s/he attended the GP for a consultation within the last two weeks.

The correlation coefficient for the relationship between being given a prescription and attendance for a GP consultation is a positive relationship, at 0.44.

Table 14: Bivariate probit model $P = f_1(X_1, \varepsilon_1)$

Dep.Var. = Individual has been given a prescription by the doctor for personal use within the last two weeks; Yes=1; No=0

	Mean (SE)		Z-value
Charge	-0.0791	(0.0710)	-1.11
Social eligible	-0.3178	(0.3704)	-0.86
Health eligible	-0.1264	(0.5998)	-0.21
Income	-0.0004	(0.0038)	0.11
QF_2	-0.0949	(0.1142)	-0.83
QF_3	0.0772	(0.1069)	0.72
QF_4	-0.0470	(0.2138)	-0.22
Employment_2	0.0604	(0.3502)	0.17
Employment_3	0.2514	(0.1945)	1.29
Ethnic_2	0.0565	(0.1913)	0.30
Ethnic_3	0.1505	(0.1820)	0.83
Ethnic_4	0.0356	(0.3420)	0.10
Marital_2	-0.0990	(0.1021)	-0.97
Marital_3	-0.1613	(0.2204)	-0.73
Age	-0.0032	(0.0070)	-0.46
Sex_1	-0.0924	(0.0070)	-0.28
Socsupport_2	-0.0771	(0.1060)	-0.73
Socsupport_3	0.0087	(0.1353)	0.06
Constant	-1.2655	(1.9201)	-0.66

Censored obs	10898
Uncensored obs	1903
Wald chi2(18)	28.71
Log likelihood	-5803.539
Prob > chi2	0.0521

Rho	0.4390	(1.7819)
	LR test of indep. eqns. (rho = 0):	
Chi2(1)	0.22	
Prob > chi2	0.6362	

* 5% level of significance; # 10% level of significance

The correlation coefficient, as stated previously is 0.44 which states that there is a positive 'random' error correlation – this suggests that 44% of the random error (white noise) can be explained by the relationship between an individual attending the GP for a consultation in the last two weeks and having received a prescription for medicine in the last two weeks.

The socio-economic factors reported above show that there is no significant impact upon prescription writing. It would seem that the level of prescription charge interestingly does not have a significant impact upon whether the GP writes a prescription given that the individual is a positive GP attendee.

Stage Three

This model examines the relationship of prescription dispensation with prescription writing taking into account socio-economic factors. This model attempts to grasp 'stage three' of the GP consultation and prescription use relationship previously explained. In this section we look at the 'will the individual use the prescription if s/he is written a prescription by the GP'. Of course, at this stage the data used to capture this relationship is the 'were you given a prescription in the last two weeks' and whether the individual is currently on prescription medication, of course this leads to some issues regarding if a patient is written a prescription and is already on medication then we may not grasp the use and writing prescription relationship, also it may be that a patient may have been given prescription in the last two weeks, however the course is

now finished and s/he is not currently on medication. Another effect, though possibly negligible, is when the patient is given a prescription to be used only when the need arises.

Table 15 contains the results for the sample selection model, where only those individuals who have been written a prescription in the last two weeks are examined. Table 16 reports the results for all individuals, with the dependent variable being whether they have been given a prescription in the last two weeks or not, given positive GP attendance within the last two weeks.

As we can see that in Table 15, age is a significant factor in determining if the patient is currently on medication conditional upon being written a prescription in the last two weeks, whilst in Table 16 employment status – economically inactive and gender - male are significant in determining whether the individual has been given a prescription in the last two weeks conditional upon being a positive GP attendee.

The correlation coefficient for the relationship between currently being on prescription medication and being written a prescription medication in the last two weeks is a negative relationship, at -0.56.

Table 15: Bivariate probit model – sample selection equation $M = f_1(X_1, \varepsilon_1)$

Dep.Var. = Individual is on prescription medication; Yes=1; No=0			
	Individual is on prescription medication if s/he was written a prescription by the GP in last two weeks		
	Mean (SE)		Z-values
Charge	0.3969	(0.5701)	0.70
Social Eligibility	2.4235	(3.3971)	0.71
Health Eligibility	2.5609	(3.4550)	0.74
Income	-0.0080	(0.0071)	-1.12
QF_2	-0.1950	(0.2709)	-0.72
QF_3	0.0754	(0.2684)	0.28
QF_4	-0.1734	(0.4409)	-0.39
Employment_2	-0.4164	(0.4806)	-0.87
Employment_3	-0.0189	(0.2891)	-0.07
Ethnic_2	-0.2185	(0.3637)	-0.60
Ethnic_3	-0.3536	(0.2482)	-1.42
Ethnic_4	-0.2046	(0.7013)	-0.29
Marital_2	-0.5494	(0.3981)	-1.38
Marital_3	-0.3404	(0.4174)	-0.82
Age	0.0313	(0.0160)	1.95#
Sex_1	-0.1206	(0.2576)	-0.47
Socsupport_2	0.3264	(0.2407)	1.36
Socsupport_3	-0.0194	(0.2339)	-0.08
Constant	-1.2784	(4.0946)	-0.31
Censored obs	12599		
Uncensored obs	202		
Wald chi2(18)	15.45		
Log likelihood	-1111.793		
Prob > chi2	0.6309		
Rho	-0.5622	(0.6519)	
	LR test of indep. eqns. (rho = 0):		
chi2(1)	0.08		
Prob > chi2	0.7774		

*5% level of significance; # 10% level of significance

Table 16: Bivariate probit sample selection model $P = f_2(X_2, Y, \varepsilon_1)$

Dep.Var. = Individual has been written a prescription by the GP within the last two weeks, given that s/he attended for a GP consultation in the last two weeks; Yes=1; No=0			
	Individual written a prescription by the GP in the last two weeks, given that s/he attended for a GP consult		
	Mean (SE)		Z-value
Charge	-0.0795	(0.0509)	-1.56
Social Eligibility	-0.2555	(0.3056)	-0.84
Health Eligibility	0.0296	(0.2773)	0.11
Income	-0.0007	(0.0018)	-0.37
QF_2	-0.0742	(0.0731)	-1.02
QF_3	0.0638	(0.0835)	0.76
QF_4	-0.0696	(0.1396)	-0.50
Employment_2	0.0071	(0.1427)	0.05
Employment_3	0.2685	(0.0725)	3.70*
Ethnic_2	0.0131	(0.1208)	0.11
Ethnic_3	0.0932	(0.0836)	1.11
Ethnic_4	-0.0178	(0.2386)	-0.07
Marital_2	-0.0857	(0.0793)	-1.08
Marital_3	-0.1027	(0.1076)	-0.95
Age	-0.0018	(0.0033)	-0.55
Sex_1	-0.1584	(0.0624)	-2.54*
Tenure_2	-0.0427	(0.0723)	-0.59
Socsupport_2	-0.0555	(0.0699)	-0.79
Socsupport_3	0.0292	(0.0812)	0.36
Constant	-1.6102	(0.3317)	-4.85*

*5% level of significance

The correlation coefficient, as stated previously is -0.56 which states that there is a negative 'random' error correlation – suggesting that 56% of the random error (white noise) can be explained by the relationship between patients being on medication if s/he has been written prescription by the GP in the last two weeks.

Age is a positively significant factor, with individuals being ~3% more likely to currently be on prescription medication given that they have been written a prescription (post GP attendance) within the last two weeks.

Both employment status (economically inactive) and gender (male) are significant in determining whether the individual has been given a prescription (post GP attendance) in the last two weeks. Economically inactive individuals are 26% more likely to have been written a prescription (if they attended the GP) in the last two weeks; males are also 16% less likely to have been written a prescription (if they attended the GP) in the last two weeks.

Prescription charge is not a significant factor within this relationship though it is interesting to see that an increase in the charge suggests a positive impact upon whether an individual is currently on prescription medication conditional upon having been written a prescription in the last two weeks. Eligibility also suggests a positive impact upon whether the individual is on prescription medication or not, suggesting a positive impact upon dispensation though this factor is not significant.

4.7 Conclusion

In this paper we have attempted to look at GP consultations as a screening tool for the government, allowing policy makers a tool to screen for potential health problems earlier thus reducing potential future cost. We hypothesised that a potential barrier to individual's attending the GP for a consultation could be the prescription charge that individuals may face, especially if they are ineligible, based upon certain health and income and factors.

The individual has two decisions to make 1) go to the GP and 2) use prescription or not, whereas the GP has one decision – write a prescription or not. Unfortunately, this paper has not been able to tackle the issue of risk information, where an individual is faced with the risk of gaining information that will lead to the decision of using a prescription or not.

Previous literature reported an effect (and some differentiated this effect between exempt and non-exempt individuals) upon prescription utilisation brought on by charges, either their introduction or increase. Since both Ryan (1989) and O'Brien (1989) reported a difference in the effect of charges, upon exempt and non-exempt individuals, with a decrease in the prescription rate of the non-exempt but no change (or even increase) in the exempt population, we assumed that given that there is no effect of the prescription rate upon the exempt population, GPs are not changing their prescribing habits but, patients are either refraining from GP contact or they are simply not dispensing prescriptions written.

The human capital framework links the relationship between the (initial) GP consultation and (final) prescription dispensation appropriately. The individual will chose to invest in his/her health and that investment based upon the initial investment (the GP contact) and, the possible investment (the prescription), thus giving the GP attendance and prescription a contingent commodity, linked, relationship. The individual's attendance for a GP consultation is contingent upon the individual's perception of receiving a prescription, especially if we assume that all individuals going to the GP (who require a prescription) will be written one, since once the prescription is received a decision to use or not needs to be made.

The models analysed within this paper are based upon cross-sectional data that is not longitudinal. However the analysis attempts to force the data into a longitudinal framework. This is a criticism of the analysis and one that can not be rectified unless further longitudinal panel type data is collected. Also another issue with the empirical analysis is that the data collected are point estimates, all the data is collected at a single point in each year (for two weeks only). Combining data from different years may also create bias in the analysis, especially where there has been a change in public perception an beliefs; however this criticism may not be valid in this instance as the data are combined over a shorter time period (ie four years).

Even though the data used within the model are not as robust as would have been preferred, the results of the analysis do fill an important gap within the knowledge of the impact of prescription charges within the health economy.

We found that though prescription charge increases and eligibility (based on social reasons) can affect GP attendance; these factors neither affect GP's decision to write a prescription nor the individual's decision to use the prescription.

We know from previous work by Ryan (1989) and O'Brien (1989) prescription charges lead to a decline in prescription utilisation, however whether this decline is a result of reduced GP attendance, reduced prescription writing or, reduced prescription use was unclear. In our analysis we examine the impact of prescription charging and prescription charge increases upon the GP –patient pathway as a three stage process. Stage one – individual's decision to go to the GP; stage two – GP's decision to write a prescription and; stage three – individual's decision to use the prescription. The analysis was carried out separately for patients eligible and ineligible for free prescriptions, as prescription charges (and their increases) will have an impact for ineligible patients but, due to the nature of their status, eligible patients will not be affected by prescription charges.

The prescription charge increase is a significant factor in deciding whether the patient attends the GP or not – with the greater charge leading to a decline in likely attendance; eligibility is an important factor with individuals whose eligibility is based upon employment status or income more likely to attend the GP. Males and unemployed individuals are less significantly less likely to attend the GP, whilst individuals who are economically inactive are more likely to

attend the GP. Income is a significant factor, with increased income suggesting less likely attendance.

Our analysis of stage two of the GP-patient pathway suggests that if the patient attends the GP then there are no significant factors suggesting a greater or lesser likelihood of being written a prescription by the GP, however it would seem that increased charge and positive eligibility status have a negative impact upon prescription writing, though the impact is not significant.

At stage three of the pathway, use or not conditional on them being given a prescription in the last two weeks (and given positive GP attendance), male patients are significantly less likely to dispense their prescriptions; individuals with an economically inactive status are more likely to dispense their prescription. The prescription charge or eligibility does not have a significant impact upon use though all three have a positive coefficient value.

From all the empirical analyses conducted it is viable to conclude that prescription charge increase in itself does not significantly impact upon prescriptions used or prescription writing. The relationship between GP attendance and prescription writing is a positive one (if you attend the GP and need a prescription you will be written one), whilst the prescription writing and use relationship is negative (if you are written a prescription after attending the GP then you are less likely to use) – this is interesting, and may be caused by the 'are you on a medication now?' proxy that was used to assess the dispensation likelihood, the limitation of this proxy is that individuals may have been given

medication for a shorter period of time (ie 7 days) thus though they dispensed their medication they are not on medication now – possibly causing bias. Our analysis would suggest that though prescription charges may deter the individual from attending the GP for a consultation once, they are not a sufficient barrier to prescription writing or use.

From our analysis we were able to conclude that the GP consultation is an effective method to allow individuals to attend for a type of screening though there is possibility they may be deterred by a prescription charge.

4.8 References

Grossman, Micheal., *The Demand for Health*, NBER, 1972.

Hardman, RA., "The Effects of Abolishing Prescription Charges," *The Practitioner*, 1965, 351-7.

Lavers, RJ., "Prescription charges, the demand for prescriptions and morbidity," *Applied Economics*, 1989, 1043-52.

Leck, Ian., "Effects of Prescription Charges on Medical Care.," *British Journal of Preventive and Social Medicine*, 1972, 56.

O'Brien, B., "The effect of patient charges on the utilisation of prescription medicines," *Journal of Health Economics*, 1989, 109-32.

Ryan, Mandy., *Estimating the effects of prescription charges on the use of NHS prescribed drugs in England, 1979-1985*, Aberdeen: 1989.

5 The extent to which the decision-making process is affected by the methodology: Health economic modelling tools

5.1 Introduction

Health economic modelling is becoming more and more important in assisting with resource allocation decisions within the health care field, especially with the setting up of national bodies such as the National Institute of Health and Clinical Excellence (NICE), All Wales Medicines Strategy Group (AWMSG) and the Scottish Medicines Committee (SMC) all of which assist the Department of Health in advising local decision makers about access to certain health care resources.

Today we live in a society where resources are constrained[29], especially in the health care environment where budgets are constrained; in such situations economic modelling is required, in order to maximise expected utility and to inform the decision maker of the optimum decision.

In the ideal world one would know the costs and the outcomes of introducing a health technology onto the market and a simple calculation could be performed to assess its impact. However, we live in a non-ideal world with imperfect information about probabilities of events occurring, possible outcomes and, costs attached to these events; there is also a lot of uncertainty surrounding individuals' behaviour, disease pathways, and the effect of the technology. The decision maker needs to maximise welfare and minimise costs to society, this is

[29] Petitti (2000)

where economic modelling is most valuable; the methodology can be used to fill the 'gap' in the decision-makers tool-belt, the modelling process allows the available data to be extrapolated out to inform decisions and allow forecasts to be made.

Modelling is used by decision-makers to facilitate appropriate resource allocation, and in order to provide an equitable health service to all (Claxton et al. (2000)); since 1999 the National Institute of Health and Clinical Excellence (NICE) has published 87 health technology appraisals (National Institute of Clinical Excellence (2005)), including cost-effectiveness analyses.

Modelling allows the individual to extrapolate cost and effectiveness data beyond that observed within a clinical trial or gained from literature reviews or medical experts, thus allowing the modeller and decision-makers to view the impact of introducing a new technology without relying heavily on expert (gu)estimates or needing large life-time clinical trials to see what might or might not happen. Modelling also allows the modeller to link intermediate clinical end-points with final health outcomes, such as a decline in blood pressure by (for instance) 5 mmHg could lead to a decreased risk of stroke and thus translate into an improvement in quality adjusted life years (QALYs), so producing a meaningful outcome.

Health economic modelling takes places when a resource allocation decision needs to be resolved and the current data are either inadequate or inapplicable, for instance when clinical data are available for 12 week period but the impact of

the health intervention is likely to be over a 10 year period. Within health economic modelling an intervention, be it pharmaceutical, interventional or methodological, as long as it has definable outcomes and inputs, can be modelled. Even when 'hard' (clinical trial) data is not available modelling is still a very important tool, since in such cases the data required to feed the model is gathered from a 'softer' source – such as expert opinion (Buxton et al. (1997)), which though not ideal is a valid information resource.

It is important to explicitly identify all the variables relevant to the question (ie is the new health intervention cost-effective in comparison to current therapy) being asked when undertaking a health economic modelling exercise (Pidd (2003)). One needs to identify the model structure, the relevant disease states, the options (the interventions under study – the 'new' therapy and it's competing alterative therapies), the controllable variables (where the modeller can define the passage that the patient will take (decision nodes)), the uncontrollable variables (where the pathway of the patient will be decided by chance (chance nodes)), and the data requirements (Pidd (2003) Barton et al. (2002)). Below the three modelling techniques available to the health economic modeller are set out with such a framework in mind.

Three modelling methodologies are available; the decision tree, the Markov model and discrete event simulation. This paper examines the structure and data requirements of each model, the similarities and, the differences, as well as the (dis-) benefits of applying one specific methodology to a problem. In conclusion

creating a framework by which to judge when a specific methodology would be most ideal to use.

Section two of the paper discusses the need for modelling within health economics; section three describes each of the three techniques and compares them. The fourth section illustrates the application of these techniques by a practical comparison. The final section concludes the analysis finding that the choice of model is very heavily influenced by the specific question to be answered (the remit of the modelling exercise), the data available, timeframe of the problem, the intended audience and, of course modeller choice. The actual result produced by each methodology will be of similar magnitude and direction, if the same data and assumptions are applied across each model-type.

5.2 Economic Models

5.2.1 What modelling techniques do we have access to?

Decision analysis, the term used to encompass the models available, is a logical quantitative method used to assess the relative value of a number of different decision options. Decision analysis allows decision makers to make decisions regarding patients' management or treatment, based upon the option that provides the 'best' outcome or the most value (Petitti (2000)). In other words, modelling provides the decision-maker with all the relevant information required to make an informed decision about treatment.

Within decision analysis there are three distinct types of models – Markov modelling, decision trees and discrete event simulation, of which discrete event simulation is the least used. A review[30], carried out in 2000, found that of the 119 economic evaluations published in 1997 74 (62%) employed decision trees, 43 (36%) used Markov processes and only 2 (2%) used discrete event simulation. Each of the three models has different data requirements, is used to model over a different time period, and varying levels of complexity.

To produce a valid and meaningful decision analytic model a number of processes need to be undertaken; these are identified by Petitti 2000[31] as:

1) identify and bound the problem
2) structure the problem (involves construction of the decision analysis)

[30] Karnon (2003)
[31] Petitti (2000)

3) populate the decision analysis with the appropriate information required

4) analyse the decision analysis

5) finally carry out a sensitivity analysis.

In the sections below each of the decision analytic models is created step-by-step in a theoretical framework, using the process mentioned above and additional ones required by the specific model.

5.2.1.1 Decision Tree

Decision tree modelling is one of the simplest decision analysis methodologies available in the modelling arena, it is ideal for modelling uncomplicated scenarios. Using the steps identified above we set up the theoretical decision tree, with a practical example.

The first stage – identification and bounding – involves a number of processes, beginning with the identification of the alternative courses of action available, the other components of the problem (events that occur following the treatment, such as a complication) and the final outcome.

An illustrated example could be with type II diabetes; where a new treatment (pioglitazone) has been discovered and the impact of its introduction needs to be assessed; the alternative course of action would be the metformin-gliclazide combination (this is a moot model as pioglitazone has already be appraised by NICE and guidelines upon it's use have been published). A possible complication with type II diabetes is that the patient develops nephropathy (decreased kidney function); there are many other complications of diabetes, but

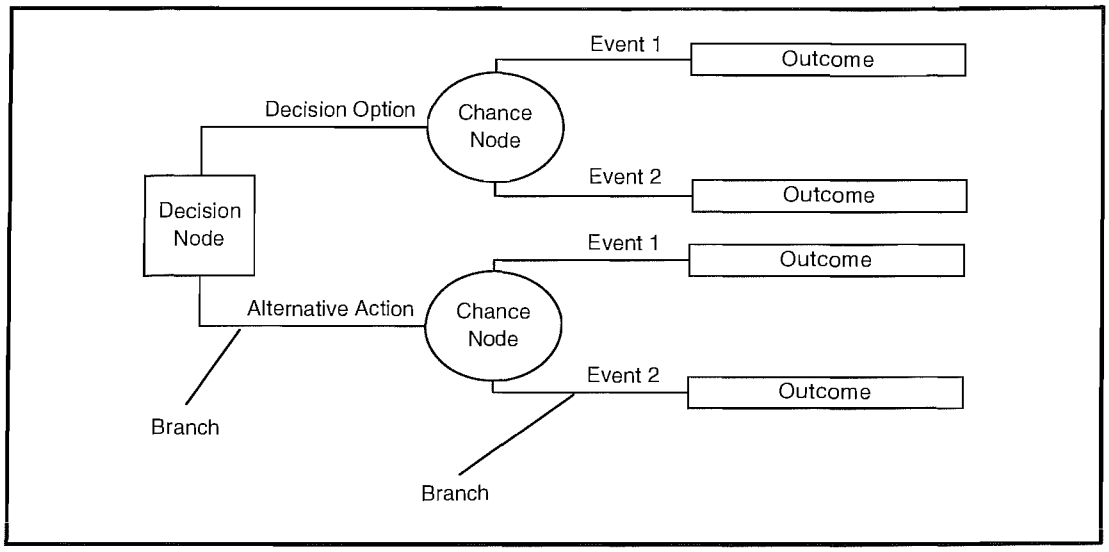
this is used for simplification. The final outcome identified in our example is kidney transplantation; this is a possible outcome though one would hope that health care professionals would intervene well before this becomes a reality.

Therefore with the introduction of a newer treatment (in this case pioglitazone) into the treatment gambit for type II diabetes we have, in this stage, identified the comparator (metformin-gliclazide combination), possible components of the problem (risk of nephropathy and the risk of kidney transplantation), and the final outcome of kidney transplantation. This of course is a highly simplified view of type II diabetes; however, this is simply for the purposes of clarity.

The second stage – structuring the problem - requires the problem to be constructed as a decision tree, which is a graphical depiction of the decision problem and relates consequences to actions. To build a decision tree, one must start building from left to right, when time is an issue earlier events are placed on the left and later events on the right.

The decision tree is made-up of decision and chance nodes, branches and outcomes. By convention decision nodes are depicted as squares, chance nodes are circles and outcomes as large rectangles. Branches connect nodes to nodes and nodes to outcomes, and are usually drawn at right angles to nodes (Figure 6).

Figure 6: Hypothetical Decision Tree Petitti 2000[32]



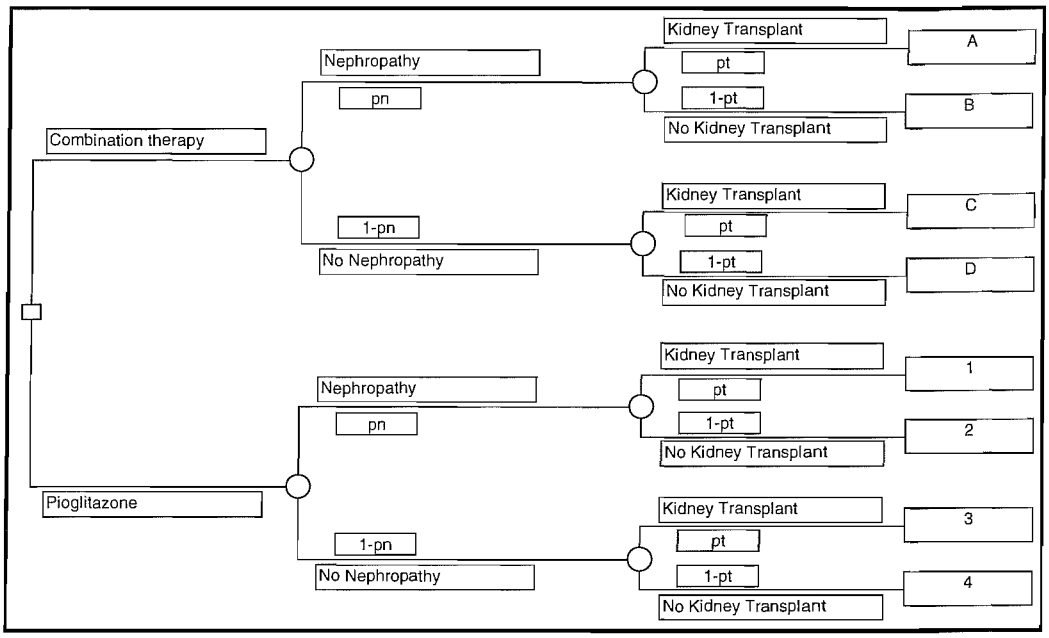
Points at which the decision to take an alternative action is dependent upon the decision-maker are decision nodes (in our simple diabetes example this would be where we decide between placing the patient(s) on the metformin-gliclazide combination treatment or pioglitazone).

Chance nodes represent those points where events beyond the control of the decision-maker may occur. Probabilities are associated with the chance nodes, and at any chance node the sum of the probabilities is one. In our simplified diabetes example, the chance node would be where the patient faces a probability of developing nephropathy or not, followed by a second chance node where the patient has a probability of requiring kidney transplantation or not, this is process is depicted diagrammatically in Figure 7.

[32] Petitti (2000) – p.19

A diagrammatic example of the above decision tree is given below:

Figure 7: Diagrammatic view of the pioglitazone vs combination therapy decision tree



Outcomes are the consequences of the final events depicted in the tree, in our example this could be Kidney Transplantation versus No Kidney Transplantation. Outcomes can depict life/death, extension in life, quality of life, or quality adjusted life years (QALYs), it would depend upon the problem under review, and the data that are available.

The third stage – populating the model with appropriate information required – requires the modeller to gather together probability information for each chance event. This will be carried out by conducting literature reviews (including meta-analysis), gathering primary data (from clinical trials), and/or consulting with experts. The model is then populated with the relevant information.

In our the simple diabetes example we would need probability information on the chance of developing nephropathy (p_n) or not ($1-p_n$) and the probability of requiring a kidney transplantation (p_t) or not ($1-p_t$), the kidney transplantation probabilities depend upon whether the individual had nephropathy or not. Probability data could be gathered from published clinical diabetes literature and/or clinical experts. The patients' flow through the model is uni-directional and the probabilities at each node are dependent upon previous states, thus the probability of nephropathy is dependent upon the therapy arm (pioglitazone or metformin-gliclazide combination), whilst the probability of kidney transplant is dependent upon the therapy arms and whether the patient developed nephropathy or not.

The fourth stage – analysis of the decision analysis – is carried out by a process called folding back and averaging, to calculate a final probability estimate of the expected outcome of each decision alternative. One would need to calculate the expected probability of each possible outcome in each alternative arm (i.e. by multiplying the probabilities together) then sum together the expected probabilities of arms representing the same outcome to establish a final expected probability for the decision option.

In our example this would mean the probability of having a kidney transplant if you are a type II diabetic would be the probability of a kidney transplant on metformin-gliclazide combination plus the probability of a kidney transplant on the pioglitazone therapy arm. This mean the modeller needs to make an implicit assumption that the decision to treat the patient with pioglitazone or combination

therapy would produce a probability of the patient developing nephropathy (this probability could be acquired from the clinical trial literature available, guidelines, epidemiology or expert opinion), then the patient would have a probability of requiring a kidney transplant and this would be dependent upon whether the patient has nephropathy or not (again data is gathered from the clinical literature, epidemiology or clinical experts). The outcome/cost at each end node is representative of the costs/outcomes associated with the specific pathway. In our example this would mean:

Combination Arm			
Probability of Nephropathy	**Probability of Outcome**	**Outcome**	
p_n	p_t	Kidney transplant	A
p_n	$1 - p_t$	No kidney transplant	B
$1 - p_n$	p_t	Kidney transplant	C
$1 - p_n$	$1 - p_t$	No kidney transplant	D

Expected probability of kidney transplant = A+C

Expected probability of no kidney transplant = B+D

Therefore, the probability of an individual requiring a kidney transplant when being treated by combination therapy is the sum of the probability of needing a kidney transplant if the patient has nephropathy and the probability of needing a kidney transplant if the patient does not have nephropathy.

Pioglitazone Arm			
Probability of Nephropathy	**Probability of Outcome**	**Outcome**	
p_n	p_t	Kidney transplant	1
p_n	$1 - p_t$	No kidney transplant	2
$1 - p_n$	p_t	Kidney transplant	3
$1 - p_n$	$1 - p_t$	No kidney transplant	4

Expected probability of kidney transplant = 1+3

Expected probability of no kidney transplant = 2+4

In order to compare the two strategies, the results of the expected probability calculations are subtracted from one another; so in order to calculate the effect of pioglitazone versus combination therapy on the expected probability of kidney transplantation we can say this is (A+C) – (1+3).

The fifth and final stage – sensitivity analysis – is carried out to assess the overall stability of the analysis conclusion. The sensitivity analysis will vary the probabilities one at a time (or together if required) while holding other variables constant.

In the model outcomes will be attached to the 'end' nodes whilst costs will be attached to the 'events' within the tree as they occur, as well as to the 'end' nodes. To calculate the outcomes the patients will pass through the tree and will be assigned their respective outcomes at the end node. To calculate the total costs, the costs for each intervention are accumulated along each pathway

(dependent obviously on probabilities at the chance nodes – where the direction of the patient is decided) and summed at the end.

Decision trees, which are the simplest form of decision analysis, are the most commonly used economic modelling techniques. However, one issue with the decision tree is that it is not possible to, explicitly, model time within the model; time can only be accounted for by the outcome measure. At each end-point, the model can only account for an average time between events (represented by chance node) for each unique pathway through the model. The decision tree model can not model how long (in our example) each individual has to nephropathy and then from nephropathy to kidney transplant.

Decision trees are the most relevant model type to use when events are to occur over a short time period, or where evaluations have an intermediate outcome measure, such as screening programmes, however if the number of events to be modelled increases, or the complexity of the problem increases, the decision tree can become large, cumbersome and highly complex. Essentially the model is a series of joint probabilities resulting from an exogenous decision to place the patient upon one treatment or another. It is not possible to tailor the joint probabilities upon other factors such as age, gender, previous medical history for each patient. It is also not possible to allow the length of time an individual has been in a specific health state affect the probability of movement to another state. Another weakness of the model is that time and events that occur over a longer period of time can not be factored into the model; it is not possible to create a

decision tree model that runs over ten years with a probability of an event occurring every year for the ten year period, the model would just be too big!

5.2.1.2 Markov Modelling

The Markov modelling process is a complex method of modelling an intervention that involves a number of different transitions (or events), where the impact of the intervention will occur (or has effects that occur) over a longer period of time more than once. The Markov model does not rely upon joint probabilities (unlike the decision tree, where you can only go to a given state based on where you have been), the probability that an individual will move from one state to another is dependent upon the state s/he is in and the treatment s/he is being given. The Markov model allows the modeller to create a model that is over a longer time period and unlike the decision tree allows the modeller to allow time as a factor into the process, ie probabilities can affect an individual at certain time points repeatedly, and these probabilities can change over time, too.

The Markov modelling process tries to provide an accurate representation of the complex processes that require patients to move in to and out of various health states; the risks and probabilities associated with these states can change over time (Petitti (2000)). Markov models are most appropriate when used to analyse conditions where patients flow through set disease states and where there is a need to consider long time periods.

In a Markov (chain) model events are modelled as transitions from one state to the next. The time horizon of the model can be split into different cycles, all being of an equal length. At the end of each cycle the patient either moves into

another subsequent health state or remains in the same state. This process of state transition carries on until all patients move into an all-absorbing state, such as 'dead' or until the 'set' model length is reached (ie five or 10 years).

Each cycle length is dependent upon the analyst and should be set to reflect clinical appropriateness. The movement from one state to the next is based upon transitional probabilities. Transitional probabilities are conditional probabilities (conditional upon the current health state), transitions will occur every cycle until the 'end' state is reached by all patients or the 'set' model length is reached. The transitional probability of an event occurring may (or may not) vary with each cycle (ie time dependent transitional probabilities) however the time point at which the event occurs will remain set throughout the model

There are another four steps identified by Petitti 2000[33], in the Markov process that are required, as well as the stages identified earlier, these 'extra' stages fit into the 'identify, bound and structure the problem, and populate and analyse the decision analyses' sections:

1) Identify the health states and describe the transition methods (identify, bound and structure)
2) Determine cycle length (structure/populate)
3) Determine the transitional probabilities (populate)
4) Estimate the outcome with and without intervention (analyse)

[33] Petitti (2000)

Using the stages identified above by Pettiti and the diabetes example illustrated previously we set-up an example Markov model.

In our over simplified type II diabetes example we can say that there are two interventions, metformin-gliclazide combination and pioglitazone; the patient is currently a managed type 2 diabetic and has a probability of remaining so (that is without complications) or has a probability of developing nephropathy, developing end-stage renal failure, having kidney transplantation, or dying.

The first stage – choosing the states and transitions – is carried out by initially depicting the Markov process graphically, where the states are by convention defined as ovals or circles. Arrows that link one state symbol with another are used to represent the allowed transitions between the states of the model.

In our example that would mean that our states are defined as managed type II diabetes (without complications), nephropathy, end-stage renal failure, kidney transplantation, and death. The transitions (or movement) allowed in the model are from managed type 2 diabetes (without complications) to nephropathy or death, nephropathy to end-stage renal failure or death, end-stage renal failure to kidney transplant or to death, and kidney transplant to death, once in the death state the patient would remain there.

Figure 8: Hypothetical Markov Model Type 2 diabetes

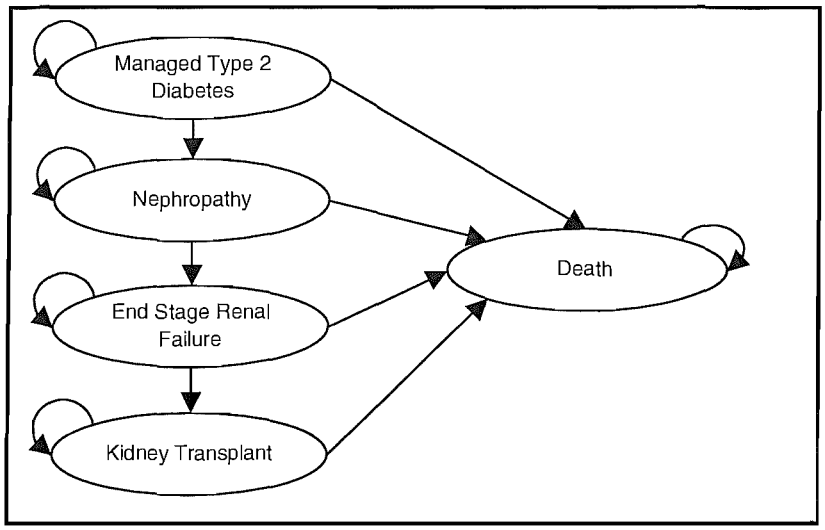


Figure 8 depicts the Markov process graphically, using the type II diabetes example, showing progress through the model from one state to the next.

The second stage – choosing the cycle length – requires the modeller to decide the amount of time that elapses between a possible movement from one state to the next. The cycle length is chosen to represent the underlying biological process that is being modelled, it may be short (weeks) or long (years); the cycle length is identical throughout the model. The cycle length is based on clinical evidence gathered from literature or expert consensus, and is set to be neither so long that it does not capture transitions (ie misses them) nor so short that there is an added complexity that is not required.

In our type II diabetes example the cycle length could be set to five years if this, based on literature and expert consensus, is felt to be of an appropriate length to capture fully the transitions possible.

The third stage – determination of transitional probabilities – is a very critical part of the modelling process and would usually take the form of data extraction from the literature, if available or, from expert guidance. If transitional probabilities are not available or are not calculable then the state transition, no matter how important for the disease, can not be included.

Most available information about transitions between clinical states is expressed in the form of a rate *'r'* (i.e. the number of events per unit time), which then can be converted to a transitional probability *'p'* of the event occurring over a time interval *'t'* based on the following formula: $p = 1 - e^{-rt}$ (Petitti (2000)) where *'e'* is the base of the natural logarithm. This is different to the probabilities in the decision tree methodology where the probabilities are not time related, the decision tree probabilities are conditional probabilities that is the patient's progress through the model is dependent upon where the patient came from, however progress through the Markov model is only dependent upon the state the patient is in.

Fourth stage – estimation of outcome – can be carried out with three main methods: monte-carlo simulation, analysis of hypothetical cohort (Markov cohort simulation) or, matrix algebra.

There are two orders of monte-carlo simulation – the 1st and 2nd order. With monte-carlo simulation of the 1st order individuals are followed through the model and costs and outcomes are recorded as per individual (i.e. 1st order

simulation assesses the within-patient variability), the path through the model is determined not only by the transition probability matrix but, also the output of a random number generator (thus random variation will depict path variation – the randomness is attached to the individual). Monte-carlo simulation of the 2nd order follows individuals through the model using alternative values of variables sampling from a distribution imposed upon the data in the simulation and costs and outcomes are recorded per individual (i.e. 2nd order simulation assess the parameter variability), it is monte-carlo simulation of the second-order that is equivalent to probabilistic sensitivity analysis, model parameters are randomly sampled from the distributions imposed upon the variables (the randomness is attached to the transition probability) in the model and the influence of these variables upon final outcomes is assessed.

Cohort analysis runs a large hypothetical cohort of individuals through the model. Fixed proportions will appear in different states, and there is no measure of variability in costs and outcomes. The hypothetical cohort will be re-allocated, in each cycle, with respect to the transition matrix. Multiplying the benefits and costs of each state by the fraction of the cohort occupying that state and summing across the states would give the total benefit and total cost generated by the cohort for each cycle. The cumulative total is the running total of benefits and costs generated during each cycle, and will tend towards a limiting value, since the cohort will move into the final state (usually death) as the each cycle progresses. Sensitivity analysis to assess outcome sensitivity to parameter variability is also possible.

Matrix algebra is a simple mathematical method of coming to a solution where the transitional probabilities are static, without the use of simulations. The monte-carlo simulation and cohort analysis methods can be adapted to carry out a sensitivity analysis, as described above.

The Markov model can be applied to a prevalent, incident or mixed cohort of patients. The costs and outcomes (utilities, quality of life etc) can be attached to each health state modelled, outcomes can be attached to reflect the severity of each state, and the costs are attached to associate the cost of remaining in *a state* for *a single cycle*.

For the model to be a true Markov model the experiences or history of patients in each disease state must be independent of how they arrived in that disease state, time with the disease and how long they have been in a specific disease state can not affect progression though the model; however these rules may be broken and this is not uncommon.

Markov chain models are particularly advantageous for modelling where events occur over a long period of time (a lifetime model examining impact of different health interventions upon the risk of a cardio-vascular event), whereas the decision tree is advantageous for shorter time periods (a one year model examining the impact of a pain therapy upon chronic pain management in a terminally ill cancer patient). However the limitation of the modelling process is that a single time period must be chosen after which patients may move into the next state (cycle length), so in our diabetes example after five years individuals

are subject to a transition probability where they could move to a different state, it is not possible to vary (increase or decrease) the length of time after which individuals face the transitions again. However it may be possible (though a little messy and awkward to handle) to link different Markov chain models using different cycle lengths to represent different parts of the patient's possible disease progression, ie when the patient is in the diagnosis model the cycles are five yearly but, once the patient moves into nephropathy model the cycles are yearly. The other limitation with Markov is that transition probabilities are not influenced by the pathway taken to arrive at a given health state, as can be done in the decision tree model. This could be remedied by splitting health states (i.e. (A|A) in state A prior to state A where an individual faces an 'x' probability of moving to state A in cycle 'y' from state A in cycle 'y-1' if s/he is was in state A in cycle 'y-2'; and (B|A) in state B prior to state A where an individual faces an 'z' probability of moving to state A in cycle 'y' from state A in cycle 'y-1' if s/he is was in state B in cycle 'y-2' and so on), but as can be appreciated this increases model complexity, adds dimensions to the number of TPs required and may make analysis cumbersome.

5.2.1.3 Discrete Event Simulation

The discrete event simulation (DES) modelling process is probably the most complex and data hungry modelling technique; this process is used when modelling required is for a population, where patients of different ages, gender, ethnicity, etc, are modelled and each of these factors could have a possible effect upon the treatment and outcome of the model. Although one could try to recreate the same process with a Markov model this would require having a different set of transitional probabilities for each different characteristic

(different transitional probabilities for each characteristic within a range as well as combinations of characteristics – ie male aged 30-40 years old will have different transitional probabilities than a female age d 30-40 years old), thus requiring a large number of transitional probability vectors and creating a large cumbersome model. The DES techniques is also required when outcomes and costs are viewed over the longer term or lots of short cycles, in which case decision trees may be too cumbersome.

In the DES technique patients move through the model experiencing events (based on a probability) at any discrete time period after the previous event, so unlike the Markov process the time period where a transition occurs can vary, either independently or be assigned by the modeller, this is also different to the decision tree – the decision tree (when accounting for time) has transitions occurring at set points so that the overall time at the end of the model is set and does not alter.

In our type II diabetes example we could (if the data allowed) say that a patient aged 30-40 years old, male, diagnosed with type II diabetes for 10 years had a 'x' probability of nephropathy after five years of treatment with pioglitazone whilst a male aged 40-50 years with only a five year diagnosis of type II diabetes and five years treatment with pioglitazone has a 'z' probability of nephropathy, where 'x' is not equal to 'z'. Therefore if we use the decision tree diagram from before (Figure 7) we can see that the 30-40 year old patient will move through the model at a different rate facing different probabilities to the 40-50 year old patient.

Patients can be assigned attributes such as age, or stage of disease before they enter the model, they may also acquire attributes as they move through certain events in the model, such as time with disease, vaccination or, previous illness. The attributes of a patient will influence their pathway through the model, as well as the costs and outcomes associated with the pathway travelled. DES enables the modeller to allow attributes (such as individual characteristics and time with disease) to affect the patient's pathway. If the decision tree were to do this the decision tree (as can be anticipated) would be incredibly large and complicated.

With this method of modelling one individual can be followed through at a time, allowing sampling from a probability distribution, and as said before allowing different types of patients (eg of varying ages) to be passed through. One patient is sent through the model at a time, each patient will face a different probability of an event occurring based upon factors such as previous history or side effects (toxicity from cancer treatments). The time point at which an event will occur can also vary i.e. probability of the event occurring straight after the previous event, in one month's time, three months time and so on. Thus in our diabetes example the male type II diabetic patient aged 30-40 years old who develops nephropathy may develop end-stage renal failure after six months, whilst the male type II diabetic aged 40-50 who develops nephropathy may develop end-stage renal failure after two years.

DES does not have memory limitations, as experienced by Markov modelling. The transition to another state can be influenced by the patient's state history as

well the patient's individual attributes. DES allows a greater flexibility, allows a more 'realistic' modelling approach, which given appropriate data, can lead to greater confidence in the results produced. However, as is common, reliable data may not be available, therefore DES can run 'what-if' scenarios, which allow the modeller to examine the impact of certain factors upon end results in the form of a sensitivity analysis.

The DES model, with it ability to allow attributes to influence probabilities and thus pathways, needs only to be run once to incorporate all these different attributes, whilst a Markov model would need to be run a number of times with attributes being altered for each run.

Although DES allows greater flexibility, this may be at the expense of requiring greater specialist analytical input, which could cause problems in model validation (since there is an increase in complexity); there may also be a greater requirement of modeller time and costs (of acquiring data) in order to create the model. The model may become over-specified where pathways are made more complex than necessary, leading to greater data requirements.

The DES does seem to be the 'ideal' model, ie can take account of differences within a population that affect each patient's transition through the model, however the DES model is very data hungry and so complex, thus though the ideal method to model a population as true to real life as possible DES is not in the majority of cases creatable.

5.2.2 How do the standard models differ to one another?

There are two factors that can assess the appropriateness of the economic model to be used, flexibility (the ease of adaptation to different forms of data and parameter reality – i.e. how close to reality does the model apply) and analytic input (the time and complexity required to develop the model and the ease of modification). The following sections compare decision tree, Markov model and DES methodologies to assess the similarities and differences.

5.2.2.1 Decision Tree vs. Markov model

The decision tree and Markov model are very different methodologically and practically.

The decision tree is a model for the shorter duration, examining outcomes and events that occur over the short term, or that do not have too many complications (for example the diabetes example where patients are on either metformin-gliclazide combination or a pioglitazone and will either have nephropathy or not, is an incredibly simplified example of the true life disease progression of a type II diabetic patient).

The Markov model – allows the disease to be modelled with greater complexity and over a greater time period (the diabetes example – can be modelled with increased microalbuminuria (for a number of years), leading to possible nephropathy, followed by possible remission or progression and so on).

Creating a decision tree model that allowed the additional complexities as described the Markov diabetes example would be incredibly complicated.

However, it is also important to note that the Markov model can not model events based upon history, unlike the decision tree – the Markov model has no memory and so the progression to the next state is independent of how you arrived at the current state, the decision tree allows pre-dependence.

5.2.2.2 Markov vs DES

Both Markov modelling and DES can be used to model a disease pathway over an extended time horizon.

With the DES model, patients will move through the model experiencing probabilities of events occurring at any discrete time point after the previous event (with both probabilities and timing dependent upon the individual's characteristics). The analysis of the model is triggered by the occurrence of an event, at which the model will ask itself what happens next.

The Markov model moves patients through the model experiencing probabilities of events occurring at set time periods, the analysis of the model is triggered by the set time period, the model will ask what events are occurring at regular events, rather than what next (the DES); the probabilities and timing of events for each individual is fixed.

Within the DES model it is possible to apply transitional probabilities related to, for instance the age, gender and, time since diagnosis, of an individual. However, if this was required within a Markov state then there would need to be a number of state-entry dependent transitional probabilities (other than start state) required – such as patients, for example, aged between 50 and 55 years (5 states),

male and female (2 states), with between 2 and 4 years (2 states) time since diagnosis would require – $(5^2)^2 = 625$ states, creating huge modelling complexity.

In our diabetes example, where patients are treated with either metformin-gliclazide combination or pioglitazone therapy, the Markov model allows the disease to be modelled with a certain level of complexity – states such as increased microalbuminuria, nephropathy, kidney failure, kidney transplantation, and death are included in the model – however these events (states) will occur at set time points. With the DES model these same events will occur however, they will not occur at set time points; unlike the Markov model the patients in the DES model will move through the model experiencing an event at discrete time points (since timing of event is dependent on individual characteristics), thus leading to a more realistic patient pathway.

The DES model also allows the modeller to take account of history; if a patient has increased microalbuminuria for a number of years, then in the DES model s/he will have a greater probability of nephropathy which can be taken into account however, this historical element can not be modelled into the Markov model.

An examination of the differences/similarities and the positives/negatives of the Markov vs DES model has recently been carried out; Karnon (2003) examined the use of alternative economic modelling techniques, for combination therapy

(tamoxifen and chemotherapy) versus tamoxifen alone in node positive, postmenopausal women aged under 65.

Applying the same disease pathway and similar assumptions and data Karnon (2003) simulated both the DES and Markov models. The outcomes reported from both models were very similar with cost differences under £500 and outcome differences under 0.15 quality-adjusted life-years (QALYs) for both models, resulting in a very small difference in the incremental cost-effectiveness ratios (ICERs – the extra cost required for one unit of benefit gained) of £3483 for DES and £3365 for Markov models (2.5^{th} percentile: DES - £452, Markov - £588; 95^{th} percentile Tamoxifen dominates – which means that Tamoxifen is cost saving as well as having greater benefit gain than Tamoxifen and chemotherapy); although the mean difference in costs, in QALYs and in life years are different for both models the direction of these results is the same.

Further analysis of the results shows that in the CEAc (cost effectiveness analysis curve) the tamoxifen and chemotherapy arm in the Markov model has a slightly greater probability of producing positive net benefits for all values of an additional QALY than the DES model. Even though both models produced very similar results there were differences between the models. One difference was survival time, these were used as available in the DES model but had to be altered (or transformed) to constant transition probabilities in the Markov model. Karnon found that the transitional probabilities and set survival times went against the Markov model whilst the DES technique allowed a more flexible approach given available data. However Karnon concludes that the closeness of

the results would suggest that it is unlikely that the use of one model over the other would lead to an alternative resource allocation decision. In such a situation, as the described by Karnon, the DES is more flexible but given the greater time required in developing and evaluating the DES the Markov maybe more preferable.

The DES model and the Markov model are complex models, with the DES having the extra complexity of allowing the modeller to account for history and allowing the patient to experience events at discrete time points. The Markov model is less complex model not allowing patient history to be taken into account, and events occur at set time points however, the Markov model technique is less data hungry than the DES modelling technique. It would seem that the DES model is only beneficial when the modeller holds strong prior beliefs regarding the data and potential influence on the end result.

5.2.2.3 Decision Tree vs. DES

Decision tree and DES are the two most similar, yet most different, of the modelling techniques available to health economic modellers. Similar to the decision tree the DES model allows history to affect the patient's progression through the model; however unlike the decision tree the DES allows modelling over a longer time period and of a heterogeneous population to occur in a simpler fashion.

The main difference between the DES technique and the decision tree is that with DES the modeller is allowed to model 'different' types of patients (be they different based upon age, gender or other criteria) through the model at the same

time, allowing a form of population modelling to occur; the decision tree is unable to carry out this function at one time, it is feasible to pass different types of patients through the decision tree model but one would have to 'run' the model a number of time (ie as many times as there are different patients) to fulfil the same 'population' criteria as the DES model.

DES is a complex model that allows the modelling of patients with respect to characteristics that can influence the probabilities and progression faced by the individual when moving through the model this is essentially the main difference between decision trees and DES.

Another difference between the models is the 'data' reliance of DES; decision trees are (as said before) the simplest of the three modelling techniques – and are possibly the least data hungry; the DES technique – most probably due to the population and historical allowance – the most data hungry.

If we take the simple diabetes example it obvious to see that the decision tree would not allow the same level of complexity as the DES model would; allowing events to occur at discrete time points is not a capability in the decision tree (it is in DES), longer term models are not possible in the decision tree as this increases complexity.

It is very easy to confuse, or even to consider, that the decision tree and DES are one and the same, especially when the decision tree incorporates the capability of passing one individual through at a time – allowing each individual to receive a

different probability of an event occurring (monte-carlo simulation of the 1st order). This is similar to DES in that individuals are passed through the model separately and each individual faces a different probability. However it is important to realise that the distinct difference is that with the decision tree the individuals are the 'same' as in the same range of probabilities are be assigned because there is no difference in the age/gender or other diagnostic factors that could effect probabilities; in the DES analysis, patients may well (as they are passing through the model) acquire different attributes – they may not be allowed to pass down a certain therapy because they've already had that therapy or they are at increased risk of an event due to things that have happened in the past, thus can pass through the model with different pathways, and each individual can face a different probability based upon these 'factors'.

Above the differences and similarities between the three modelling methodologies have been examined, there may be some fairly obvious reasons as to why, for instance, decision tree modelling instead of Markov modelling or DES is the model of choice, such as model remit (the question to be answered), the time frame and complexity of disease; however one of the underlying reasons for the choice of model is based heavily upon data – availability and limitations.

Data availability is a highly important factor when deciding the type of model to create, an example of this would be new intervention 'P' developed with highly detailed clinical data requiring a cost-effectiveness model to be created for the purposes of reimbursement; however the alternative therapy 'M' has incredibly poor clinical data. The clinical data for new therapy 'P' it would suggest the

ideal model would be to create a Markov model however given the poor comparator data the actual model developed is a decision tree. Similarly, with detailed clinical trial data available for 'P' and assuming that there is a specific difference in effect of 'P' on different patient types (for instance related to years since diagnosis) it could be possible to create a DES model however given the lack of comparable comparator 'M' data this type of model is not possible.

Of course, choice of model is not only limited by the lack or availability of the data but also by the 'question' that the model is trying to answer. It maybe, though the occurrences of this are likely to be rare, that the analyst has a choice of models from decision tree to Markov through to DES model, however if the remit of the model is a simplistic budget impact model (where only the impact upon the health service of introducing a new technology is assessed) then a simplistic decision tree model would be sufficient, whereas a model that would be supporting part of a health technology submission such as NICE (National Institute of Health and Clinical Excellence) or the SMC (Scottish Medicines Consortium) would benefit from a more complex modelling process such as Markov or DES modelling.

5.3 Practical Example

As mentioned above, the modelling methodology used to resolve a given health technology is dependent upon the type of disease area being studied, the data available, time span of the model, comparators to be included in the model, and modeller choice!

It is possible that the choice of modelling technique used may (or may not) give rise to a different result being achieved. The possible differences/similarities produced by the use of different modelling methodologies is examined in the paper Karnon (2003) where the technology under analysis, alternative adjuvant therapies for early breast cancer, was assessed by using two methodologies Markov modelling and discrete event simulation.

This paper will attempt to contribute to the growing literature and knowledge in this period by further extending and examining the decision tree versus discrete event and the decision tree versus Markov modelling processes, thus developing a guide and tool to the selection and impact of one modelling methodology versus another, as well as bringing forth the potential hurdles to overcome when using a specific methodology.

As previously discussed there are three types of methodologies that can be used in the modelling process, the decision tree, Markov model and discrete event simulation, in the following section we examine the impact of using each of these three methodologies in the analysis of a given technology.

The technology in question is a new therapy for moderate to severe psoriasis. Psoriasis is a complex disease that can be managed with a gambit of therapies; at the mild stage the therapies include placing the patient upon topical agents, such as E45 cream. However, a percentage of psoriasis patients are diagnosed with moderate to severe psoriasis, which is a complex disease manageable by a number of possible systemic therapies (for the purposes of this analysis they are to be labelled 'A', 'B', 'C' and 'D'). The new technology, which we will name 'R', is an alternative therapy offered to patients once they have failed all the other therapies ('A', 'B', 'C' and 'D'), and in theory have no treatment options available to them; so theoretically they would be treated with 'Z' a low-cost, no-efficacy, management therapy.

In the treatment pathway there are three methods by which the patient is able to move from one therapy to another; 1) patients have a serious adverse event causing them to stop the therapy in question, 2) patients, after achieving a satisfactory response, remain on the therapy in question for the maximum duration allowed after which they are moved and, 3) patients do not achieve a satisfactory response and are classified as non-responders; in all three instances individuals are moved to the next therapy in the pathway (Figure 9). The model length is set at five years, in order to catch appropriate patient movement through the model pathway.

The benefit, defined as response time, (patients that achieve a PASI (psoriasis severity index score) reduction greater than 75%) is only assigned to the patient

if s/he is successful on the given therapy. However, patients are not assigned any benefit if they fail (whether due to non-response or adverse events (AEs)).

The model is based upon probabilities and resource use extracted from the literature, where this data was not available assumptions based upon clinical experts have been made and are explicit within the model. Cost data is taken from published sources, such as the British National Formulary, Personal and Social Security Resources Unit (PSSRU) and the NHS Reference Costs database.

The following section will be examining the above problem within the context of three different methodologies, the decision tree (with 100 patients) methodology is compared to a discrete event type model (with 100 patient passed one at a time); the decision tree methodology is compared to a Markov model (both with 100 patients).

The models, in both comparisons, are simulated 10,000 times (using the macro capabilities available in Excel™) and the results examined, this is in order to assess the effect of the data upon the final result, to carry out a sensitivity analysis. The first comparison, decision tree vs discrete event, uses the same data and assumptions for both models; whilst the second comparison, decision tree vs Markov model, although using the same data and assumptions in both models as the first comparison does not incorporate all of the therapies in the modelling process. Different data and assumptions are applied to both comparisons in order to allow for the differences in methodologies, for instance the decision tree vs Markov comparison does not model arm 'A' due to the added

complexity that this would add to the comparison. Simplifying assumptions have been made throughout the models to allow an ease of comparison.

Table 17: The data used in both the decision tree and DES modelling methodologies

Therapies	% with PASI>=75 at point one			% of discontinuers due to AEs at point one		
	Mean	Min	Max	Mean	Min	Max
A	75	68	83	3	0	7
B	18	2	34	5	5	6
C	60	54	66	28	25	31
D	17	9	25	10	9	10
R	23	21	26	3	3	4

Table 18: The data used in both the decision tree and Markov modelling methodologies

Therapies	% with PASI>=75 at point one			% of discontinuers due to AEs at point one		
	Mean	Min	Max	Mean	Min	Max
B	18	2	34	5	5	6
C	60	54	66	28	25	31
D	17	9	25	10	9	10
R	23	21	26	3	3	4

Both Table 17 and Table 18 report the probabilities used in the modelling methodologies; these data were obtained from the literature and clinical experts. The min and max columns report the range of data that are taken from the literature and clinical experts and are used in the sensitivity analysis of the model in order to assess the degree of influence the data has on results.

The reason for choosing the psoriasis example is purely one of practicality; at the time of conducting the analyses of this paper I was involved in analysing the cost-effectiveness of various treatments of psoriasis, however the data used in this paper for one therapy was commercial-in-confidence (at the time) and for this reason was anonymised. The data used is now within the public domain.

5.3.1 One model a number of different ways

5.3.1.1 i) Decision tree vs. Discrete event

The decision tree vs. discrete event type comparison examines two pathways, R and NR.

Patients treated in the 'R' pathway will remain there for the entire model length if they respond (RP) at the initial efficacy point (based upon literature and available guidelines), however if patients fail (they either do not respond to the therapy (NRP) or experience an adverse event serious enough to warrant discontinuation (AE)) they will be moved onto the Z therapy, which means patients incur the cost of the Z therapy but they do not achieve any benefit or experience any serious adverse events; the Z therapy is the a management therapy that does not incur any benefits to the patient.

Patients treated with the NR pathway will remain on the pathway for the entire model length, moving through therapies A→ B→ C→ D → Z, and remaining upon each therapy for a one year period if responding (RP). If the patient fails at

the initial assessment point, either being non-responsive (NRP) or experiencing a serious adverse event (AE)), with a therapy (i.e. A), they would move onto the Z therapy for the remainder of the year, then at the end of the year they will move to the next therapy (i.e. B).

Costs are assigned to the patients for the entire period s/he is treated with the therapy; benefits are assigned only to the successful patient after the patient has been allowed a time to respond (Figure 10)

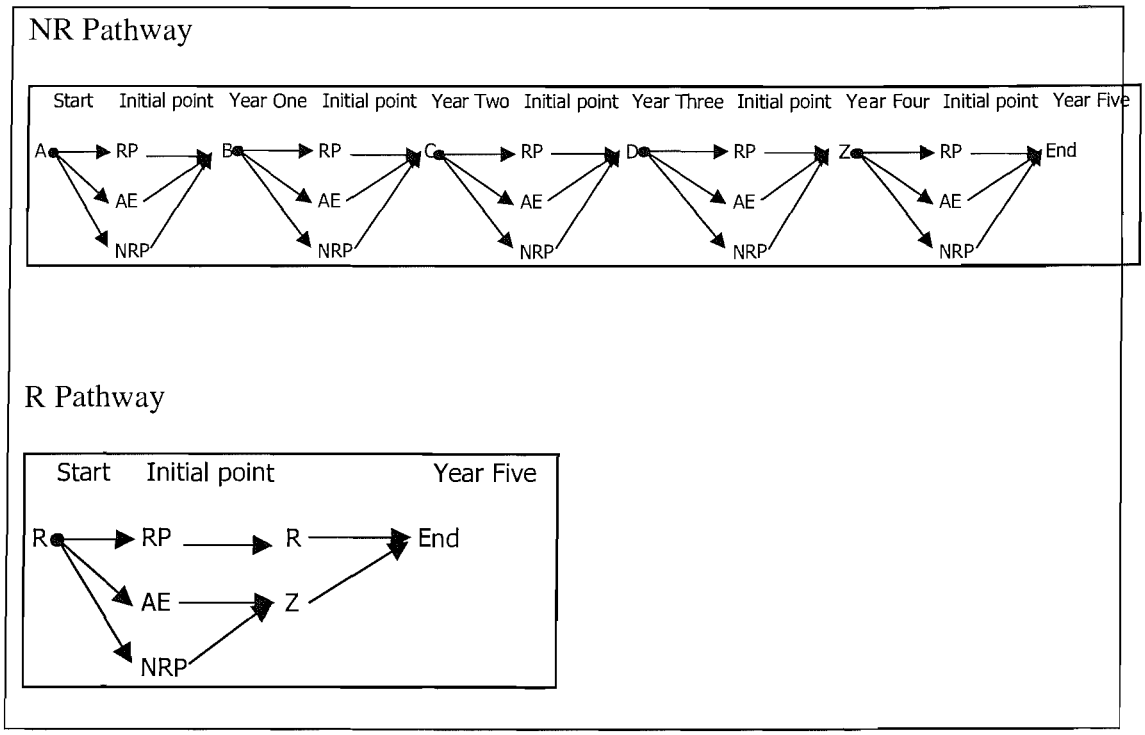


Figure 9: Graphical depiction of the two pathways (R and NR) that the patient will face in the DES and decision tree models.

The decision tree model is created and run for 100 patients; this means that 100 patients are passed through the model and will react to the mean probabilities used and are assigned the relevant costs and benefits.

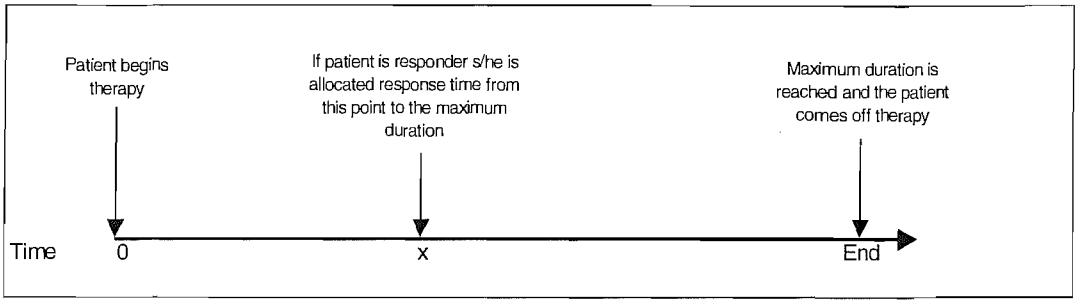


Figure 10: Graphical representation of the response time allocation

The discrete event model will pass one patient at a time through the model until 100 patients have been treated, each patient will be assigned a random number; based upon this random number the patient will be classified as a responder or a failure (an example is when the patient is being treated with 'A', s/he is assigned a random number greater than 0.75 then the patient is not responding, if the random number is less than 0.75 then s/he is a successful responder).

The DES vs decision tree difference is not only in the fact that the DES methodology will pass one patient through the model at a time instead of 100 patients at once (decision tree), but that at each chance node (where the patient faces a probability, either for response or failure (AE or non-response)) in the DES model the patient also faces a random probability to determine whether the patient responds or not, suffers an AE or not, or is a non-responder. Therefore there is a different pathway being created at each simulation (or patient run).

Unfortunately, this is not the exact definition of a discrete event simulation model; a true DES model is defined by the specific parameter-related

probabilities required. So we would need a probability of success for patients treated with therapy 'A' based upon their age, or gender or, time with disease; for instance the patient has a greater probability of being successful if the patient is female rather than male (of course there is not any data available to validate this!). However, such data are immensely difficult to come by and this is one of the main hurdles when creating a DES model.

Decision Tree			
Cost of R	**Benefit of R**	**Cost of NR**	**Benefit of NR**
£8,509.89	0.72 years	£15,227.36	1.16 years
Incremental Cost		**Incremental Benefit**	
-£6,717.48		-0.44	
ICER			
£15,126			

Discrete Event			
Cost of R	**Benefit of R**	**Cost of NR**	**Benefit of NR**
£15,604.82	0.95	£14,892.65	0.99
Incremental Cost		**Incremental Benefit**	
£712.17		-0.04	
ICER			
-£17,804			

Figure 11: Results from the decision tree vs. discrete event comparison (static results)

The results of running both models are different to each other see Figure 11 above, though, the direction of the result is similar, in that the incremental benefit of 'R' over the 'NR' pathway is negative.

The explanation for the difference could be that the discrete event model is rather erratic; with each simulation, as one individual passes through the model at a time, the individual is assigned a set of characteristics, these characteristics are taken from a uniform distribution (random number), thus with each simulation a 'proxy' individual is produced, thereby producing a different pathway (given that

each individual will react in a different manner) and thus a different set of costs and benefits. As can be visualised, the results that are produced for 100 patients, in the decision tree model, where each patient only has to consider that each chance node s/he faces has the same set of probabilities attached to it that the remaining 99 patients will also face will differ from the 100 patients, in the discrete event model where each patient will face a different random number at each chance node, than his/her fellow 99 patients, and thus this random number influences the pathway through the model. With every run of the model the discrete event model will produce a different result, whilst the decision tree model will produce the same result every single time. However the discrete event model can gravitate towards producing the same result with every run of the model – if the model is programmed to run a large number ($\geq$ 10,000 simulations).

The above simulation is a static simulation – which means that all patients (DES or decision tree) will face the same probability of an event, whilst the DES model patients also face a random factor that will determine whether (based upon static probabilities) the patient will respond or fail – below the results of a sensitivity analysis are reported in order to assess the impact of the probabilities upon the end result.

Decision Tree			
Cost of R	**Benefit of R**	**Cost of NR**	**Benefit of NR**
£8,673.93	0.74 years	£15,214.20	1.16 years
(£6,599.92, £10,801.13)	(0.48, 1.00)	(£14,758.14, £15,671.48)	(1.04, 1.28)
Incremental Cost		**Incremental Benefit**	
-£6,540.23		-0.43	
(-£8,555.15, -£4,157.05)		(-0.67, -0.18)	
ICER			
£15,343			
(£12,183, £26,133)			

Discrete Event			
Cost of R	**Benefit of R**	**Cost of NR**	**Benefit of NR**
£13,565.42	0.69	£14,878.93	0.98
(£8,000, £46,187.40)	(0.00, 4.77)	(£9,251.15, £19,505.35)	(0.00, 2.35)
Incremental Cost		**Incremental Benefit**	
-£1,313.51		-0.28	
(-£9,423.67, £29,696.38)		(-1.67, 3.96)	
ICER			
£4,633			
(£2,042.08, £10,632.75)			

Figure 12: Results from the decision tree vs. discrete event comparison (sensitivity analysis results)

The results of running both models are different to each other, see Figure 12 above, though, the direction of the result is similar, in both the decision tree and the discrete event model the 'R' therapy is more costly and less efficacious (produces less benefit) than the 'NR' pathway. The discrete event model also

has much wider 95% confidence intervals than the decision tree model for the 'R' and 'NR' pathways.

The discrete event model is very erratic, each individual has a random probability of being successful, if the patient is not successful then s/he faces a probability of experiencing an AE serious enough to warrant discontinuation however, if the patient does not respond and does not experience a serious AE then s/he will be classed as a non-responder. Along with this random factor, the discrete event model and the decision tree both, with each simulation, vary the parameters between the 95% confidence intervals, around the mean using a normal distribution (which is assumed as being representative of the data), this is the sensitivity analysis that is mentioned as part of the steps to developing a decision analysis (Pidd).

The results that are produced for 100 patients that only have to consider that each simulation will provide them with a different set of probabilities, and that they will at each time point simultaneously either have a response, AE-failure or non-response failure (decision tree), will differ from the one patient that has to not only consider the different set of probabilities generated with each simulation but, also the different random numbers generated with each simulation that will influence the result of the therapy (discrete event).

The cost effectiveness result produced by both the decision tree model and the discrete event model indicate to the decision maker that the new therapy ,'R', is neither has a higher efficacy nor a lower cost than the current therapies ('A', 'B',

'C' or 'D'); therefore the decision maker would be advised not to implement 'R' in place of therapies 'A' to 'D'. However the decision-maker also has another tool to aid the decision-making process – the cost-effectiveness plane (CE plane). The CE plane shown in Figure 13 (below) represents (plots) the results of all the simulations onto an x/y graph and allows the decision maker to view the overall result (including monte-carlo simulations where the probabilities are varied). If the majority of the plots are in the upper left quadrant the existing treatment is dominant (less costly and more effective) if the majority of simulations are in the lower right quadrant then the new treatment is dominant (less costly and more effective), if the majority of plots are in either the lower left or upper right quadrants then the decision-maker needs to make an assessment as to the level of cost-effectiveness. To do this the decision-maker can use a cost-effectiveness acceptability curve (CEAc), Figure 14, where the incremental cost-effectiveness ratio (the additional cost required for an additional unit of benefit) is plotted based upon a number of thresholds (the maximum that a decision maker is willing to pay for the additional unit of benefit) and the decision maker can see what is the probability of the new intervention being cost-effective at his/her own threshold level.

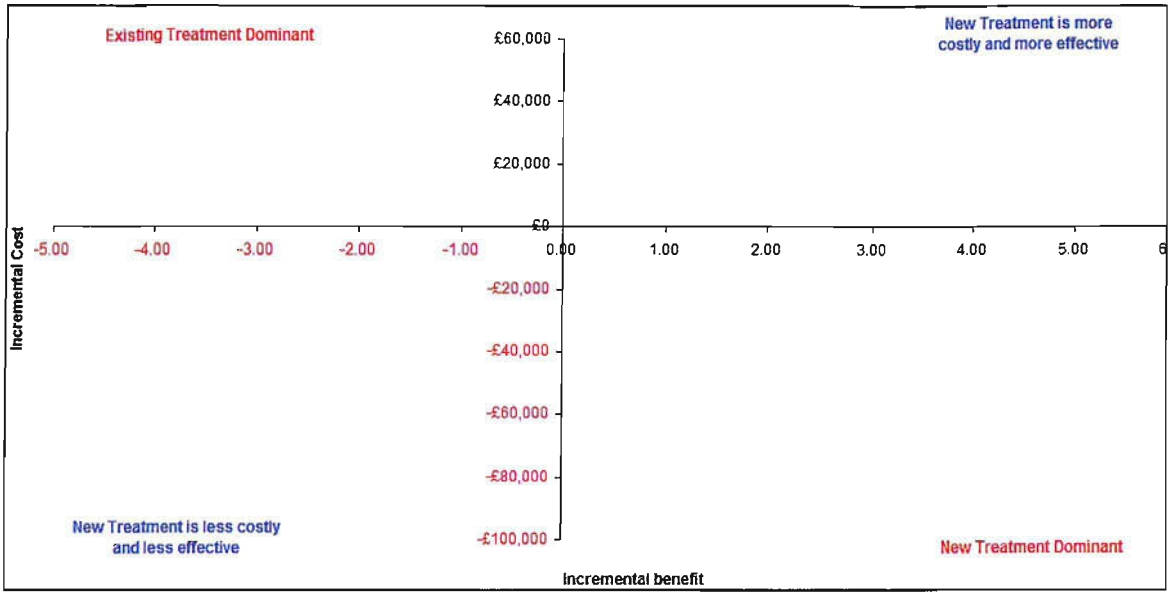


Figure 13: An example of a cost-effectiveness plane with all four quadrants showing

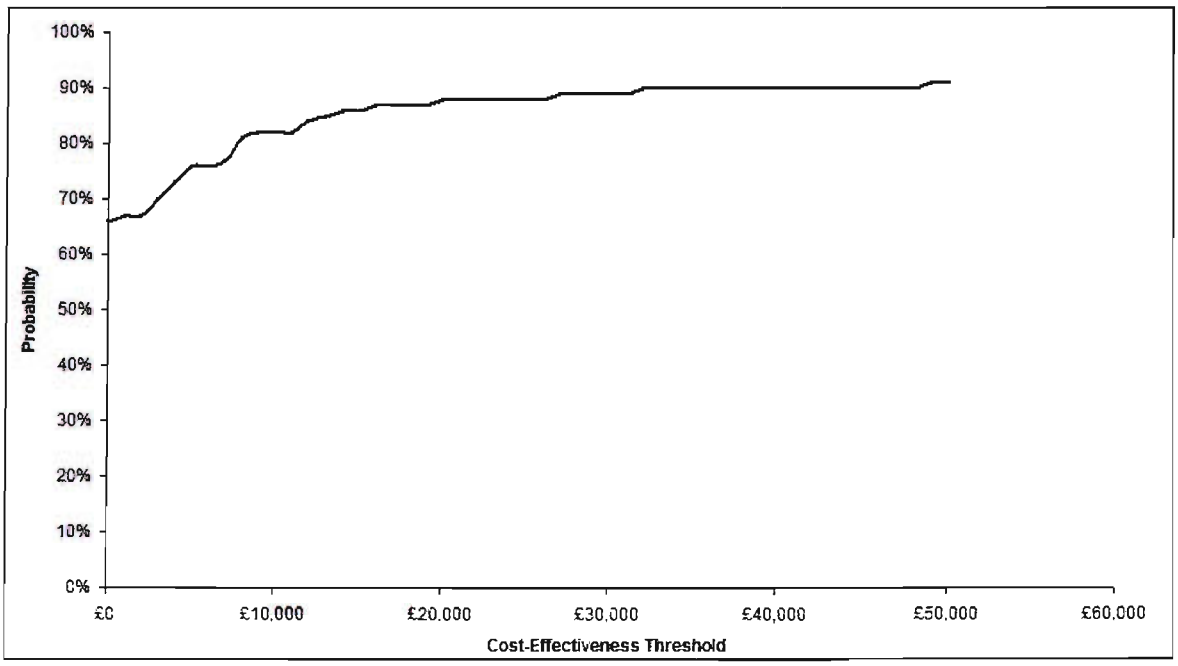


Figure 14: An example of a cost-effectiveness acceptability curve – indicating that at a threshold o £30,000 the new technology has > 85% probability of being cost-effective.

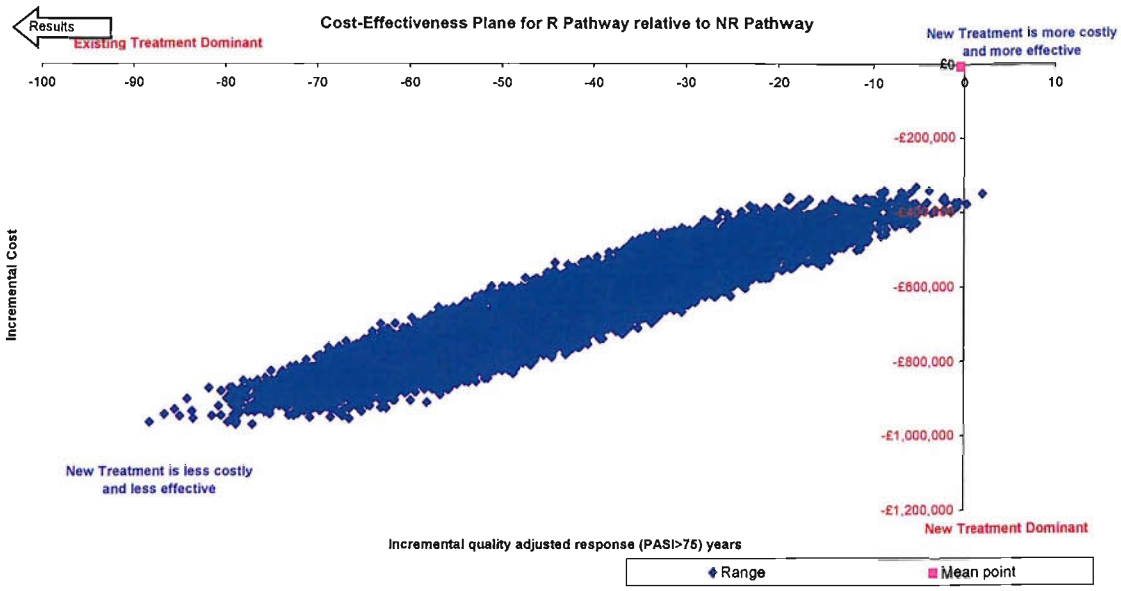


Figure 15: Cost Effectiveness Plane for decision tree model

The above figure, Figure 15, shows the cost effectiveness plane for decision tree model, this plots the costs related to the benefits of the new treatment versus the current treatments. The graph represents sensitivity analysis and shows that even with the parameter variability around the distributions the new treatment is not a cost-effective option.

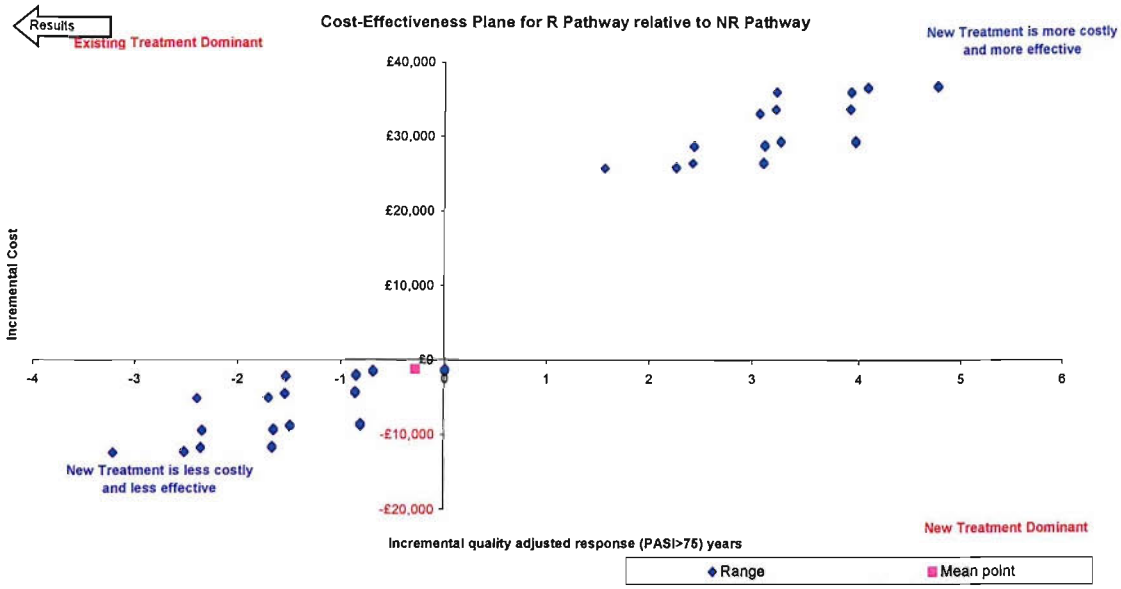


Figure 16: Cost Effectiveness Plane for Discrete Event

The figure above, Figure 16, shows the cost-effectiveness plane for the discrete event model which indicates that with the parameter variability (based upon the imposed distributions) the new treatment is either, more costly and more effective or, less costly and less effective. In such a situation a cost-effectiveness acceptability curve, Figure 17 below, shows the possible threshold levels and the probability of the new treatment being cost-effective at these levels.

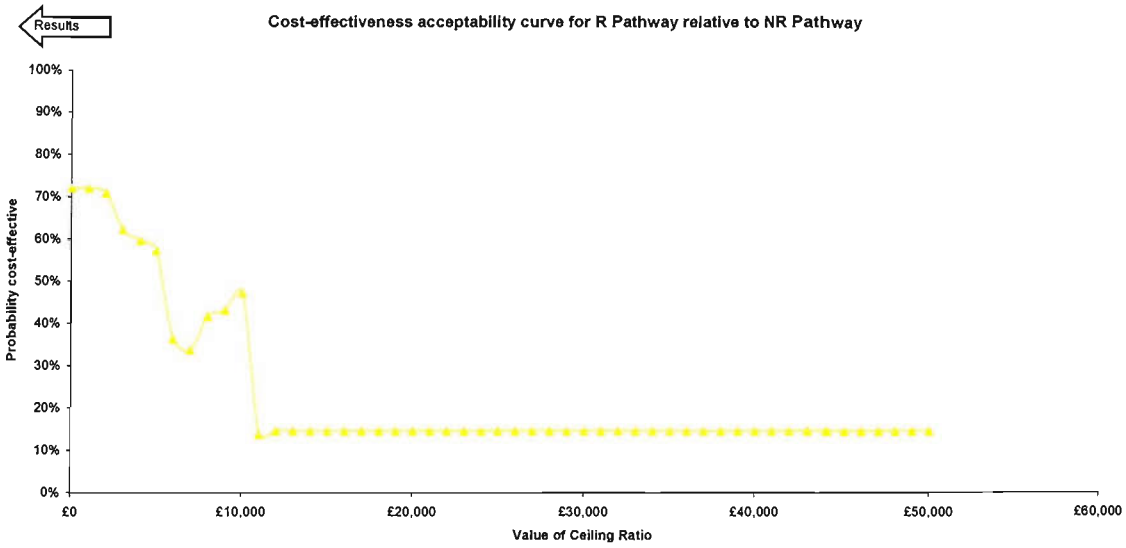


Figure 17: Cost- effectiveness acceptability curve for Discrete Event

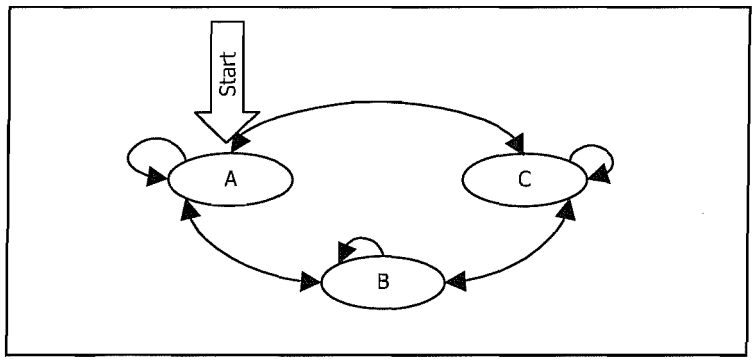
Given the stochastic nature of the patients passing through the discrete event model and the number of patients that have a negative incremental or zero incremental benefit the cost-effectiveness acceptability curve does not really provide a meaningful result. But what it does show is that the therapy has only a 10% probability of being cost-effective over the £10,000 threshold.

5.3.1.2 ii) Decision tree vs. Markov model

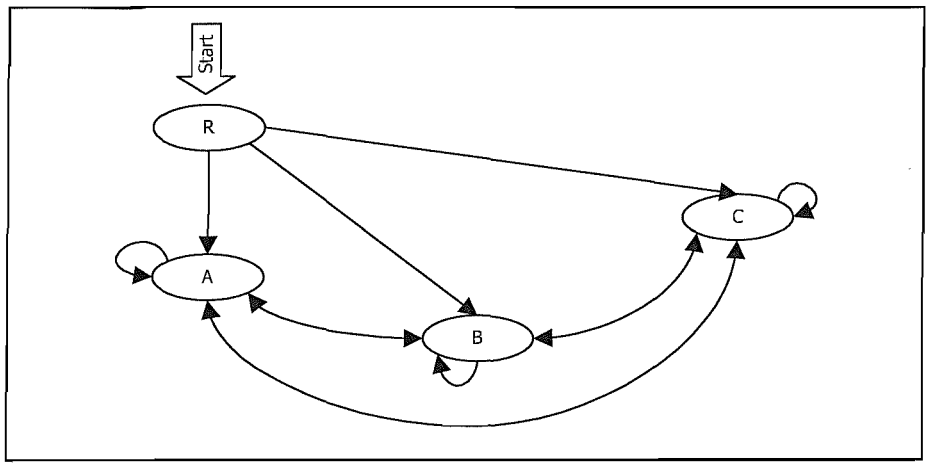
The decision tree and Markov model comparison examines two pathways, 'R' and 'NR'. Patients remain on the 'R' pathway for the entire model length, beginning with 'R' but at the end of cycle one they will face a probability of remaining upon 'R' or moving to either 'A', 'B' or 'C', this will happen at the end of every cycle (one year) until the end of the model lifetime (ie five years).

If the patient moves away from 'R' then s/he can not move back to 'R' at any point in the model. If patients fail (they do not respond to the therapy or experience an adverse event serious enough to warrant discontinuation) they will be assigned a zero cost and a zero benefit for that cycle, as it is assumed that they have no therapy. Patients on the 'NR' pathway will remain on the pathway for the entire model length, starting at 'A' but moving to either 'B' or 'C' or remaining upon 'A', each cycle period is one year long. At the end of the cycle period the patients faces a probability of moving to the next therapy or remaining on their current therapy for the next year. If the patient was to fail (either being unresponsive or experiencing a serious adverse event) upon a specific therapy (i.e. 'A') they would move to the next therapy in the next cycle (i.e. move to 'B' in cycle two from 'A' in cycle one) they would experience a zero cost and benefit for that cycle (cycle one).

NR Pathway



R pathway



Both models pass 100 patients through the model at each simulation, for the static model; the models are then run for 10,000 simulations (100 patients each), the 10,000 simulations are run to capture the sensitivity of the results.

The costs and benefits in each model are calculated in a similar fashion for each pathway. Within each pathway the patient is assigned a cost for therapy, if the patient is successful s/he will be assigned a cost and benefit for the entire cycle, whereas if the patient has failed therapy (due to AE or non-response) then s/he will be assigned no cost or benefit of therapy for the cycle at all. These are very strong simplifying assumptions however, in order to allow this type of analysis to be undertaken (ie a Markov vs decision tree comparison) without falling into a high level of complexity this was necessary.

Decision Tree			
Cost of R	**Benefit of R**	**Cost of NR**	**Benefit of NR**
£4,626.38	1.59 years	£2,604.04	1.61 years
Incremental Cost		**Incremental Benefit**	
£2,022.34		-0.03	
ICER			
-£63,941			

Markov			
Cost of R	**Benefit of R**	**Cost of NR**	**Benefit of NR**
£4,626.38	1.66	£2,604.04	1.69
Incremental Cost		**Incremental Benefit**	
£2,022.34		-0.03	
ICER			
-£63,941			

Figure 18: Results from the decision tree vs. Markov comparison (static result)

The results from the static analysis of the decision tree and Markov model are identical see Figure 18 above; whilst the results from sensitivity analysis comparison, though different are not vastly so, see Figure 19 below. The incremental cost of adopting R instead of NR is approximately £1400 greater, and the incremental benefit is approximately $1/10^{th}$ of a year less.

Decision Tree			
Cost of R	**Benefit of R**	**Cost of NR**	**Benefit of NR**
£4,046.02 (£2,765.01, £5,593.76)	1.62 years (1.43, 1.81))	£2,620.44 (£2,236.55, £3,012.29)	1.70 years (1.48, 1.93)
Incremental Cost		**Incremental Benefit**	
£1,425.57 (£172.23, £2,954.63)		-0.08 (-0.22, 0.06)	
ICER			
-£17,108 (-£117,250, £111,436)			

Markov			
Cost of R	**Benefit of R**	**Cost of NR**	**Benefit of NR**
£4,155.92 (£3,464.67, £5,161.63)	1.66 (1.60, 1.77)	£2,661.96 (£2,487.44, £2,967.74)	1.76 (1.64, 1.89)
Incremental Cost		**Incremental Benefit**	
£1,493.96 (£621.46, £2,397.25)		-0.10 (-0.21, -0.00)	
ICER			
-£14,944 (-£46,627, £37,236)			

Figure 19: Results from the decision tree vs. Markov comparison (sensitivity analysis result)

The results of the decision analysis would indicate to the decision maker that the new therapy is not a cost-effective option to employ in the place of the current therapies.

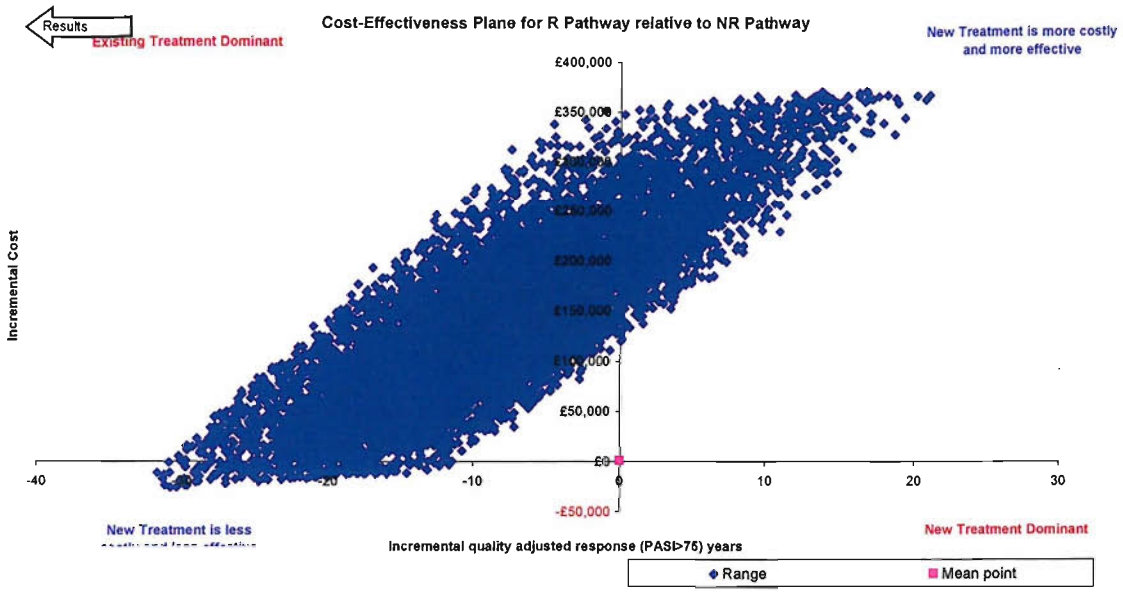


Figure 20: Cost-effectiveness plane for decision tree

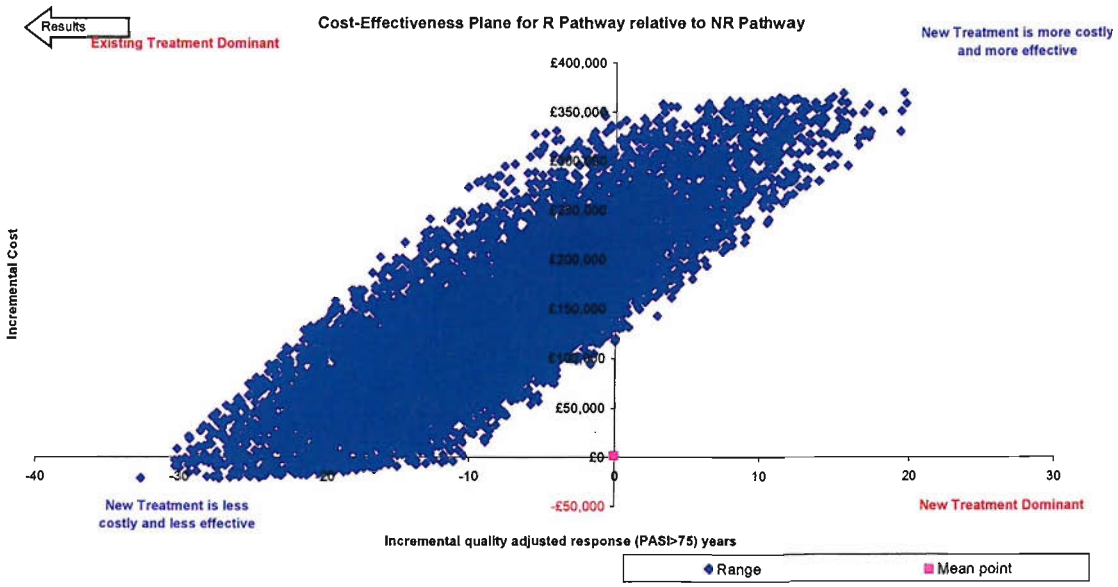


Figure 21: Cost-effectiveness plane for Markov model

Both Figure 20 and Figure 21 show that overall the new therapy is not a cost-effective treatment, with the comparator treatments dominating (ie less costly and more effective) the new therapy.

As we can see in both the decision tree and discrete event model comparison and the decision tree and Markov model comparison the direction and final results are similar and would provide the decision-maker with the same guidance. It seems that as long simplifying assumptions are made, to allow use of the same data across both models, decision tree and Markov models could be used interchangeably, however due to the inherent probabilistic nature of the discrete event type model this is probably not entirely comparable to decision tree modelling.

The DES model is the most real-life representative model, however as stated before the DES model is very data hungry, requiring parameter input data for each 'different' patient type, ie efficacy of a health technology would need to be different based upon gender, age, previous history of medications, duration of disease, co-morbidities and so on (dependent of course on whether these individual factors affected the efficacy of the health technology). However, access to this level of data is very difficult (ie it is probably not available in such fine detail) therefore it may be viable to invest additional resources to gather this additional data, in order to assess whether the additional cost is worth additional data gained; this type of assessment is an 'Expected Value of Perfect Information' analysis, where the benefit (of reducing uncertainty of parameters) is compared to the additional cost in order to inform decision-makers if this

additional information is worth the effort! However to properly discuss this type of analysis would require another thesis.

If the same data and assumptions are used, ie the data are sufficiently limited, the actual final result produced by each of the models will not cause the decision-maker to allocate resources differently. However, in order to provide the decision-maker with sufficient information to make a resource allocation decision the model must be as representative of real-life as possible, so the choice of model (and implicitly the data and assumptions that feed the model) are very important to the final result.

5.4 Conclusion

Modelling is carried out when society or individuals need to maximise their expected benefit subject to a constraint, typically within the NHS health care setting where budgets are constrained and decision makers require assistance in making decisions, such as with the NICE and SMC decision making processes.

The models available to the modeller include decision trees, Markov models and, DES. There are differences within these three models types with the decision tree being the simplest and the DES process the most complex and data hungry. The decision tree is a model for the shorter term, and examines outcomes and costs for events that occur over the shorter time frame or that do not have too many complications. The Markov model allows the disease to be modelled with a greater degree of complexity and over a greater time period however, the Markov model does not allow history to affect future events, pre-dependence is not allowed. The DES model allows the disease to be modelled over the longer term and allows history and specific patient attributes to be taken into consideration with probabilities for future events, thus allowing a more realistic model to be produced however, this means this modelling technique is very data hungry.

The practical examination of the three modelling techniques in relation to each other, decision tree vs DES and, decision tree vs Markov both carried out within this paper and the DES vs Markov model comparison carried out by Karnon (2003) all report a similarity in the direction of the results. This however, does not simply imply that all three modelling techniques are the same as one another

however it implies that the modelling techniques can produce similar results if the same assumptions are made and identical data re used. Thus if the data are sufficiently limited then the use of a more complex model will not alter the result. Methodological differences in the modelling techniques are narrowed when comparing the three techniques within the practical example, this would suggest that the modelling technique to be used is based upon not only data availability (you can not model to DES level if the data does not allow) but, also the disease being modelled (the decision tree can not fully take into account all the issues in a complex disease such as diabetes) and of course the modeller choice and intended audience for the model.

The practical comparative analysis of the three models has shown that it is difficult to force a specific problem (with a set audience, data availability and, specific issues) into one model type or another – each model type has different data, assumptions and, methodology requirements. Specifically the DES model in the practical example is not a true DES model, to be such a model probability data regarding the impact of other factors (such as disease history) would be required. Of course, this analysis was conducted upon a single case-study across the three models, and this is itself open to criticism – can one make a generalisation based upon a single case study? However Karnon (2003) conducted a similar analysis upon a very different disease area (be it only for two model comparison) however the results produced there are similar to those in this analysis. Therefore it can be argued that this result is generalisable.

This paper has found that there is no ideal technique, the only ideal technique is that suited to the problem, the intended audience and the data available. If the technology under examination is for a complex disease, over the longer term and with a specific audience remit (such as a re-imbursement process, ie NICE or the SMC) then a more complex model maybe a better technique (such as Markov modelling); however, if the technology under question has a specific effect upon the disease related to patient attributes then a DES model maybe more appropriate (given data is available); if a more simplified analysis over a shorter period or with fewer complications then a decision tree model would be most suitable.

Model-type selection is, in essence, dependent upon data, timeframe, audience and, of course modeller choice – all these variables will need to be examined thoroughly before a decision upon the model type is made.

Thus we have found that if inputs (data and assumptions) into a model structure are identical then the outputs from the model are sufficiently close to provide the decision-maker with the same guidance, regardless of the type of model to be used.

However, this does not mean that the simplest modelling methodology should always be used. As can be seen from our practical diabetes example a more complex type of model allows other important factors to be taken into account that can affect the patient's treatment pathway, disease progression, costs incurred by the health care system and of course the final outcome (in terms of

benefits and costs). Therefore, when embarking upon a modelling exercise the modeller must first bear in mind the model remit (ie the question s/he has to answer, budget impact model for a localised decision-maker or cost-effectiveness model for re-imbursement such as for NICE), the type of disease (the possible complications), the time period (necessary to capture all meaningful factors in the modelling process), the comparators required and the data available not only for the new therapy but also for the current comparators. Based upon these factors the modeller should make an informed choice with respect to the possible advantage of undertaking one modelling process over another.

A decision tree is generally more suited to a localised budget impact type model, or a short time frame with fewer complications, or where the data for the new therapy (or comparator) are limited.

A Markov model is generally more suited to a complex cost-effectiveness model, for model of a longer time frame, with a number of transitions that could occur throughout the model period but within a disease area that allows the Markovian assumption of 'no memory' to be fulfilled. However, there are now methods of overcoming this issue.

A discrete event simulation model is ideal for a longer time frame, where patient history/pathway is important but also, most important where the data is available to fill the transition probabilities for each type of individual.

5.5 References

Ashcroft, DM., Li Wan Po, A., Williams, HC. and Griffiths, CE., "Cost-effectiveness analysis of topical calcipotriol versus short-contact dithranol. In the treatment of mild to moderate plaque psoriasis," *Pharmacoeconomics*, 2000, 469-76.

Barton, Pelham, *Using Tracker varibles to overcome Markov restrictions: a cost-effectiveness model for rheumatoid arthritis*, 2002.

Barton, Pelham., Robinson, Suzanne. and Bryan, Stirling., *The use of modelling in the economic evalaluation of health acre*, 2002.

Buxton, Martin J., Drummond, Micheal F., Van Hout, Ben A., Prince, Richard L., Sheldon, Trevor A., Szucs, Thomas and Vray, Muriel, "Modelling in economic evaluation: An unavoidable fact of life," *Health Economics*, 1997, 217-27.

Claxton, Karl., Schulpher, Mark. and Drummond, Micheal., *A rational framework for decision making by the National Institute for Clinical Excellence*, 2000.

Duggan, AK., "Modelling different approaches to the management of upper gastrointestinal disease," *Pharmacoeconomics*, 1998, 25-37.

Foster, S., Godfrey-Faussett, P. and Porter, J., "Modelling the economic benefits of tuberculosis preventive therapy for people with HIV: the example of Zambia.," *AIDS*, 1997, 919-25.

Henriksson, F., "Applications of economic models in healthcare: the introduction of pioglitazone in Sweden," *Pharmacoeconomics*, 2002, 43-53.

Karnon, Jonathan, "Alternative decision modelling techniques for the evaluation of health care technologies: Markov proceses versus discrete evnt simulation," *Health Economics*, 2003, 837-49.

Knight, C., Paisley, S., Wight, J. and Jones, ML., "Economic modelling of different treatment strategies for haemophilia A with high-responding inhibitors.," *Haemophilia*, 2003, 521-40.

National Institute of Clinical Excellence, *Completed Technology Appraisals*, 2005.

Palmer, S., Sculpher, M., Phillips, Z., Robinson, M., Ginnelly, L., Bakhai, A., Packham, C., Abrams, K., Cooper, N., Alfikah, K., Flather M.; Gray, and Gray, D., "A cost effectiveness model comparing alternative management strategies for the use of Glycoprotein IIb/IIIa antagonists in non-ST elevation acute coronary syndrome.," *Health Technology Assessment*, 2003,.

Petitti, Diana B., *Meta-Analysis, Decision Analysis, and Cost-Effctiveness Analysis*, Oxford: Oxford University Press, 2000.

Pidd, Micheal, *Tools for Thinking*, West Sussex: Willey, 2003.

Richter, A., Hauber, B., Simpson, K., Mauskopf, JA. and Yin, D., "A Monte Carlo simulation for modelling outcomes of AIDS treatment regimens," *Pharmacoeconomics*, 2002, 215-24.