WILLINGNESS TO PAY FOR COLORECTAL CANCER SCREENING: A COMPARISON OF ELICITATION FORMATS.

A thesis presented for the degree of Doctor of Philosophy

at the University of Nottingham.

Emma J. Frew, BA (Hons), Msc.

2003

CONTENTS

TABLES

FIGURES

APPENDICES

APPENDICES

Appendix

ABSTRACT

Willingness to pay is increasingly being used in health technology assessment, although a number of methodological issues remain unresolved. Using data collected from four studies, this thesis presents the findings from a direct comparison between alternative format designs to elicit willingness to pay for two alternative colorectal cancer screening tests; faecal occult blood (FOB) testing and flexible sigmoidoscopy (FS) testing. Along with the willingness to pay values estimated using the open-ended, payment scale, closed-ended and iterative bidding formats, information is collected on household income, attitudes toward health promotion and personal risk perceptions to determine the nature and value of responses. In comparison with the alternative formats, the closed-ended question design produced significantly higher WTP valuations and different justifications for those valuations. It is hypothesised that the yea-saying effect may explain this difference. The payment scale format achieved a higher completion rate compared to the open-ended design and both formats produced broadly similar valuations. Although a subsequent study suggested evidence of range bias within the payment scale design. The iterative bidding format produced higher valuations than the open-ended and payment scale but lower than the closed-ended, it is hypothesised that valuations obtained using different initial bids demonstrate the existence of starting point bias. Across all studies, respondents who have a high health motivation, are well-educated, have a high household income and who are particularly worried about the disease have a positive effect on the willingness to pay for colorectal cancer screening.

Declaration

The work presented in this thesis has not been submitted in any previous application for a degree. It has been composed by the candidate, Emma J. Frew.

All quotations are distinguished by quotation marks and sources of information acknowledged.

As a result of the empirical research conducted for this thesis, the following articles have been published in peer-reviewed journals:

Chapter six

(1) Whynes, D.K., Frew, E.J., Wolstenholme, J.L. A comparison of two methods for eliciting contingent valuations of colorectal cancer screening. Journal of Health Economics; 2003, in press.

(2) Frew, E., Wolstenholme, J.L, Whynes, D.K. Mass population screening for colorectal cancer: factors influencing subjects' choice of screening test. Journal of Health Services Research and Policy; 2001; **6**; 2:85-91.

(3) Frew, E., Wolstenholme, J.L, Whynes, D.K. Willingness to pay for colorectal cancer screening. European Journal of Cancer; 2001; **37**: 1746-1751.

(4) Frew, E., Hammersley, V., Wolstenholme, J.L., Whynes, D.K. Collaborating with a primary care based research network. Journal of Evaluation in Clinical Practice; 2001; **7**; 3:309-342.

Chapter seven:

(5) Whynes, D.K., Frew, E.J., Wolstenholme, J.L. Evidence of range bias in contingent valuation payment scales; Health Economics; 2003, in press.

Chapter eight:

(6) Frew, E.J., Whynes, D.K., Wolstenholme, J.L. Eliciting willingness to pay: comparing closed-ended with open-ended and payment scale formats. Medical Decision Making; 2003; 23;**2**:150-159.

Emma Frew

Emma Frew
May 2003.

Acknowledgement

The first study that started in 1997 was originally developed by Jane L. Wolstenholme of Oxford University who helped to design the data collection instrument. I am very grateful to Jane for her constant support, advice and encouragement throughout the work done for this thesis.

A huge thanks goes to David K Whynes whose enthusiasm, support and encouragement throughout the whole research process meant the thesis got started and indeed, finished.

Thanks are given to the 'Trent Focus Research Network of General Practitioners' and to Dr Stephen Earwicker and Mr Tony Oram from 'The Health Centre' based at Stapleford, Nottingham for allowing access to patients to collect the data for the research.

I would also like to thank colleagues in the School of Economics at the University of Nottingham for their guidance and support for this thesis. These people are Paula Lorgelly, Paul Miller and Zoe Philips. Colleagues at the Trent Institute for Health Services Research and the Department of Surgery at the Queens Medical Centre, Nottingham are also thanked for their support and advice.

Colleagues elsewhere who have commented on the work gone into the thesis are Phil Shackley, Cam Donaldson, John Cairns, Mandy Ryan, Angela Robinson and Richard Smith.

Last but certainly not least I am indebted to my husband Ben and to my parents for their unlimited patience and support – thank you.

Chapter One - Introduction

The purpose of an economic evaluation is to provide advice to decision makers on health care interventions that provide good 'value for money' and therefore assist in the allocation of health care resources. Information is obtained on the costs and consequences of alternative health care strategies so that these decisions can be made. Many outcome measures exist when evaluating the effects of health care interventions. The method that is used within a cost-benefit analysis (CBA) is the contingent valuation (CV) method which is described as a 'survey based approach for eliciting consumer's monetary valuations for programme benefits for use in a CBA' (Diener, O'Brien et al. 1998). As a technique of measuring health care benefits, CV has gained popularity in the evaluation of health care technologies (Klose 1999). The main advantage of the method is that it encompasses a wider measure of benefits than the alternative Quality Adjusted Life Year (QALY) or life-year saved units that are used in cost-utility analysis (CUA) and cost-effectiveness analysis (CEA).

The CV method however does have some limitations. Considerable work is still required in establishing this approach as a reliable and valid means to assess health care benefits. There remains many methodological questions that are unanswered in the literature. The objective of this thesis is to contribute to the refinement of the method and in particular, the following questions are addressed:

(1) There remains much methodological debate in the literature as to which elicitation design (question format) is the appropriate one to apply as each has its own strengths and limitations. Within a health care context, very few studies have directly compared the performance of different elicitation methods, and among those that have, there is conflicting evidence. The primary objective of this thesis is to present evidence from four separate studies that make a direct comparison of four elicitation formats and to draw some conclusions, based on this evidence, on which is the appropriate format to adopt.

(2) Some respondents may find certain question formats to be easier to understand and complete, consequently, the response rate may vary across different designs. This project will examine the difference in response rate between the question formats and draw conclusions about the ease of completion.

(3) It has been proposed in the literature that different question formats have a direct affect on the magnitude of the willingness to pay value (Donaldson, Thomas et al. 1995). When drawing comparisons across studies that have used different question designs therefore it may well be the affect of the question format that is affecting WTP as opposed to a difference in <u>actual</u> WTP. It is therefore important to assess this difference in magnitude; this thesis will make comparisons between the WTP by question format.

(4) To assess the consistency between direction of preference and strength of preference revealed by WTP, the relationship between preference ordering and WTP can be explored. Within this thesis, participants are first asked to reveal a

direction of preference for the alternatives then asked to value using a WTP exercise allowing consistency to be checked.

(5) It is proposed that, within the payment scale question design, different ranges of values provided may have a direct impact on the average WTP value (Johannesson, Jonsson et al.1991). Two studies are presented in this thesis that use the payment scale design but with different ranges to test for this range bias effect.

(6) The iterative bidding style of questioning may be susceptible to starting point bias whereby the starting bid in the bargaining exercise influences the final WTP value (Boyle, Bishop et al., 1985). Within the study that uses the iterative bidding design presented in this thesis, different starting bids will be used to examine the impact of starting point bias.

(7) The closed-ended format may produce a 'yea-saying' effect that results in a tendency for respondents to agree to pay a bid regardless of the value of that bid (Mitchell and Carson, 1989). The closed-ended study described in this project uses different levels of bids across the sample population to explore the impact of this yea-saying effect.

(8) The choice of the appropriate instrumentation technique is discussed a great deal in the contingent valuation literature (Reardon and Pathak 1989). However the choice of instrumentation technique involves trading off the limitations and advantages of each and this thesis will discuss these trade-offs by comparing the use of mail questionnaire versus person-to-person interviews.

(9) When asking individuals to value two close substitutes it is plausible that a WTP value is revealed for the programme as opposed to the individual characteristics of each alternative. In this case, the difference in the relative WTP will be very minor and will not reflect the difference in direction of preference between the two alternatives. Throughout the thesis, WTP will be revealed for two alternatives making it important to take account of this possible embedding effect.

(10) It is interesting to explore the degree to which the WTP exercise is measuring respondents' preferences for each alternative being valued. One method of doing this is to conduct person-to-person interviews to facilitate a more lengthy discussion surrounding the reasons for value. The last study presented in the thesis presents the results of this more lengthy discussion and examines the relationship between the quantitative information (WTP) and the qualitative information.

The purpose of this introduction is to outline the structure of the thesis, in particular, the sequence in which the above issues are addressed.

Chapter two is set up as an introductory chapter with the objective of outlining the conceptual foundation of a CBA. One of the arguments for adopting a WTP exercise is that its theoretical underpinnings are rooted in welfare economics. It is important therefore to provide an explanation of these theoretical underpinnings by providing brief explanations of the concept of producer and consumer surplus along with a discussion on the revealed and stated preference approaches that exist to measures these surpluses. The

WTP method is established within the field of environmental economics therefore many of the references within chapter two stem from research conducted within the environmental field. The next chapter (chapter three) shifts the focus away from environmental onto a health care context as this will be the focus of the rest of the thesis.

Chapter three discusses the use of the CV method within health economics. The main purpose of applying a WTP exercise within health economics is to use it as a tool to measure outcome within a CBA therefore as part of an economic evaluation context. To explain the reasons why WTP may be used instead of alternative outcome measures, chapter three outlines the advantage of the method within economic evaluation. The reader is also provided with a full explanation of the purpose of economic evaluation within health care and how this information may be important to decision makers to incorporate respondents preferences into policy recommendations.

After the introduction of the CV method within health care in chapter three, chapter four presents an in depth discussion into the methodological components of the CV method. Much of the debate within health economics on the WTP method centres around the issues discussed in chapter four. This chapter is intended as a background to the empirical work that is presented in later chapters. The discussion on the methodological components within the empirical section will draw on the issues discussed in chapter four. It is designed to 'set the scene' for later discussion that will follow in the empirical section. The chapter is split into two parts:

(a) Discussion on the relationship between WTP and willingness to accept (WTA) and a brief description of the alternative elicitation formats.
(b) A thorough discussion of the various methodological components of the technique such as the validity, ordering effect, starting point bias/ yea-saying, range bias, choice of instrumentation technique, choice of elicitation format, embedding effect and then finally, strategic bias.

As the empirical work will examine the difference between preference direction and value, this chapter will also present information on the evidence of how individuals form preferences and choose between different alternatives.

Chapter five is designed to provide background knowledge to the disease area discussed in the empirical section of the thesis. In all empirical sections of the thesis, the WTP method is applied to measure the value of colorectal cancer screening. There are two alternatives for screening colorectal cancer, each with different processes of care that respondents may value differently. Therefore, to understand why individuals may prefer one test to another, the nature of the disease, the purpose of screening and the differences in the processes of screening are explained in this chapter.

The first part of the thesis (chapters two to five) provide a discussion on the theoretical aspects of the CV method alongside a discussion of the disease area to which the method will be applied. The second part of the thesis, chapters six through to nine, presents the four studies that make up the empirical section.

Overall, the objective of these chapters (6-9) is to measure a monetary estimate of the value of screening for colorectal cancer screening; to assess whether Faecal Occult Blood (FOB) testing is preferred to Flexible Sigmoidoscopy (FS) testing, i.e. a comparison of the actual modalities of screening. To ascertain whether the way in which the question is asked affects the responses given and finally, to assess the impact of the demographic factors on the absolute and relative WTP valuations for each of the screening programmes.

Chapter six presents the first piece of empirical work designed to investigate the WTP methodology. The main purpose of this chapter is to compare the performance of the two elicitation methods: open-ended versus payment scale. A number of methodological issues are raised in this chapter, in particular it focuses on:

- Response rate: the overall response rate to the questionnaires as well as response rate to the actual WTP question.
- Consistency of responses: the correlation between direction of preference and strength of preference revealed by the WTP values.
- The interpretation of zero values: the interpretation of a zero WTP value can be classed as different forms of zero depending on the nature of the response.
- The embedding effect: an exploration of the relative WTP values for the two alternatives against the direction of preference.
- Criterion validity: stated WTP values revealed through the exercise will be compared to previous research that estimated revealed WTP values.
- Use of mail questionnaires: a discussion on the form of instrumentation technique.
- Comparison of WTP with qualitative data: the relationship between quantitative and qualitative data.

Chapter seven focuses on one particular bias inherent within the payment scale elicitation format; the range bias effect. Using the results already presented in chapter six that adopted the open-ended and payment scale design, chapter seven examines the impact from varying the scale. A much smaller scale is applied and then compared to the longer scale presented in chapter six – the extent of the range bias effect is then discussed.

Willingness to pay values for colorectal cancer screening have already been analysed and presented in chapters six and seven, chapter eight presents a third study administered using a closed-ended approach to elicit WTP for colorectal cancer screening. Comparisons are then made between the results of all three studies therefore the effect on the WTP value of adopting different elicitation designs is discussed. As the guidelines within the field of environmental economics advocate the use of the closed-ended approach, this chapter assesses the consistency between these guidelines and the empirical results. In particular, this chapter focuses on:

- 'Yea-saying' effect: the tendency for respondents to say 'yes' regardless of the size of bid.
- Response rate: overall response rate as well as the response rate to the WTP question.
- Consistency of preferences: the relationship between the direction of WTP and magnitude of WTP.
- Comparison with WTP values and the qualitative information: the correlation between the quantitative and qualitative information.

Chapter nine presents a slightly different approach to estimating WTP for colorectal cancer screening. WTP is again measured for the two alternative screening programmes only this time the WTP is measured using the iterative bidding format administered in an interview setting. Using interviews to measure the WTP provides the opportunity to collect a richer data set and allows a much deeper analysis of the 'thought processes' of the individuals when estimating the WTP. The objective of this chapter therefore is to estimate WTP using the iterative bidding format and compare these values to the alternative elicitation formats and to assess the value of estimating WTP using interviews as opposed to mail questionnaires. This chapter discusses the:

- Use of personal interviews: an exploration of the influences and biases inherent within this instrumentation design.
- 'Starting point' bias: the influence of the starting bid on the final WTP value.
- Response rate: Response to the WTP question.
- Consistency of WTP values with preferences: the consistency between what individuals say they prefer and what is revealed through the WTP information.
- An examination of the qualitative data collected against the WTP values revealed: a thorough analysis of the relationship between the quantitative and qualitative information.
- The use of a novel technique to encourage respondents to consider the 'opportunity cost' of the WTP method: an analysis of the results of an exercise where respondents are asked to think of and trade products worth the WTP amount that they have elicited.

Finally, chapter ten discusses the research findings from all of the empirical work presented in the thesis. This chapter discusses all the methodological components of the method presented in chapter four in relation to the results of the empirical work. Recommendations are then discussed based on the overall findings of the thesis.

Chapter Two – Conceptual Foundation of CV method.

2.1 Introduction

The objective of this chapter is to outline the conceptual foundation of the contingent valuation method (CV). To begin, a short discussion explaining the purpose of an economic evaluation will be made with a description outlining the concepts of consumer and producer surplus. A more in depth discussion concerning the distinction between Marshallian and Hicksian consumer surplus measures will follow. When health care is offered 'free at the point of use' it is not possible to measure consumer surplus using market prices therefore the analyst has to look to other methods, these will be discussed in turn. Firstly, the revealed preference techniques are outlined, i.e. the hedonic pricing method and the travel cost method and secondly, the stated preference method is introduced.

Finally the chapter will conclude with a brief summary of the history of the contingent valuation method.

2.2 Economic Evaluation

Economic analysis is undertaken to weigh up the costs and benefits of any activity. In a world where there are scarce resources, choice involves sacrifice, that is, the alternatives foregone from the production/consumption of a good – commonly defined as the opportunity cost of a decision. In theory, rational decision making involves the choice of an item that gives the best value for money, i.e. the greatest benefit relative to its opportunity cost. Thus the primary objective of an economic evaluation is to measure the opportunity cost of a given set of resources, in other words:-

'the comparative analysis of alternative courses of action in terms of both their costs and consequences' (Drummond, O'Brien et al. 1997).

In economic terms therefore, the measurement of the desirability of a project involves the systematic analysis of all its relative alternatives.

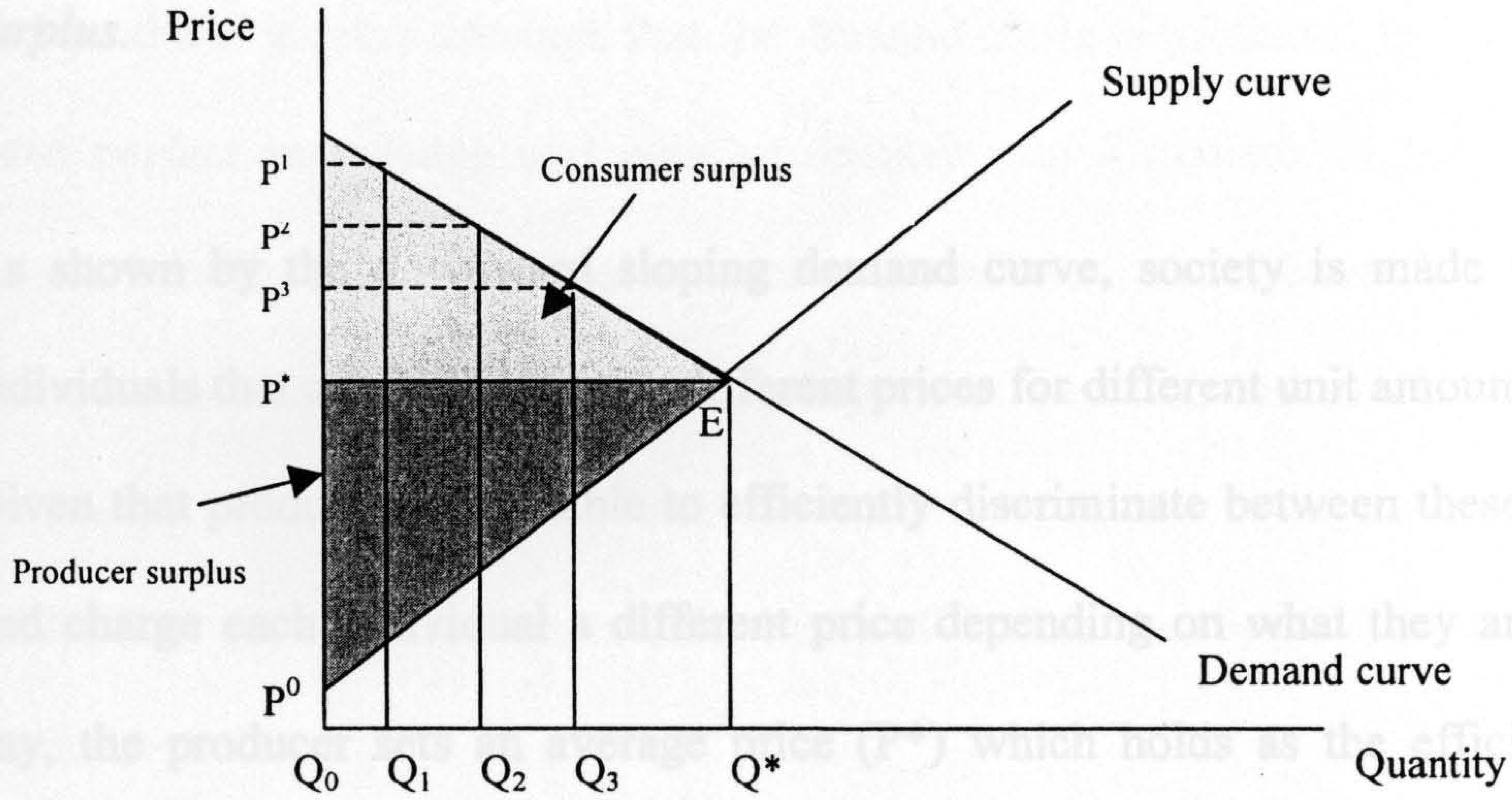
2.3 Cost-Benefit Analysis

Where the costs and consequences of a programme or project are quantified in monetary terms this is defined as cost-benefit analysis (CBA). The aim of a CBA is to measure all the true societal costs and benefits that accrue from a project so the impact is measured from the perspective of all the individuals in society whose welfare is affected. A CBA relates to social decisions about matters within the question, that affect a group of individuals (Pearce and Nash 1981).

2.4 Consumer and producer surplus

The demand curve indicates how much individuals are willing to pay for a given good in question and they are downward sloping in shape because for each additional unit, the marginal utility (level of satisfaction obtained from that extra unit) derived becomes less and less until eventually the individual is unwilling to pay for any more units of the good. A downward sloping demand curve is illustrated in figure 2.1.

Figure 2.1 Consumer and producer surplus



In figure 2.1, there is at least one individual in society who is willing to pay P^1 for one unit of the good. Similarly, there is at least one who would be willing to pay a price of P^2 for the second unit, and one who would be willing to pay P^3 for the third unit, and so on. The area under the demand curve therefore resembles the willingness to pay for the good by all individuals within the society and is the sum of all the rectangles under the demand curve.

The area, $Q_0Q^*EP^0$ below the supply curve, and above the x axis represents the actual resources it costs producers to supply Q* amount. The supplier actually receives the whole rectangle area, $Q_0Q^*EP^*$, for supplying Q* units of good, the triangle P^0EP^* above the supply curve represents a surplus above the amount that suppliers would be willing to accept. This is known as the producer surplus (dark-shaded triangle).

The benefits of consuming the good are therefore equal to the total willingness to pay (the total area under the demand curve) minus the payments required to purchase Q* units of the good. This is equal to the area below the demand curve and above the price line (lightly shaded triangle in figure 2.1). This area is known as the *consumer surplus.*

As shown by the downward sloping demand curve, society is made up of many individuals that are willing to pay different prices for different unit amounts of a good. Given that producers are unable to efficiently discriminate between these individuals and charge each individual a different price depending on what they are willing to pay, the producer sets an average price (P*) which holds as the efficient price at which producer surplus is maximised. This is what leads to a consumer surplus as there will be a segment of individuals within society that are paying P* for the good but are willing to pay more than P* as reflected by the slope of the demand curve. A change in the willingness to pay for the good will therefore represent a change in this consumer surplus.

The above theory has several assumptions associated with it. Normally, it is assumed that the market is perfectly competitive and that other markets are perfectly competitive. A perfectly competitive market assumes that there are many firms in the market that are so small relative to the whole industry that they have no power to influence price; the price therefore is determined by the interaction of demand and supply in the whole market. In a perfect market, there is also complete freedom of entry of new firms into the industry, the products produced by firms are identical in the sense that there is no branding or advertising. Finally, the perfect market assumes that consumers have perfect knowledge of the market, that is, producers are fully aware of prices, costs and market opportunities and consumers are fully aware of price, quality and availability of the product. The above analysis explaining consumer and producer surplus assumes that the demand curve is produced by consumers who have perfect knowledge and who are behaving in a manner so that the marginal private benefit to the individual is equal to the marginal social benefit of consumption. Since it is assumed that there are many other firms in the market that have no individual influence on the overall price the supply can be equated with marginal social cost (in terms of opportunity cost) of factors that are bought in competitive markets.

2.5 Marshallian and Hicksian consumer surplus

Measuring the demand for a health care technology will depend on whether its implementation or removal causes a welfare gain or loss to society.

This welfare gain or loss can be derived from changes in either:

1. Marshallian consumer surplus.
2. Hicksian (income compensated) measures.

The Marshallian consumer surplus is simply measured by the area under a Marshallian demand curve (as already described in section 2.4). The Hicksian measure is estimated from the area under a 'compensated' demand curve, where the project involves a welfare change there are two possible Hicksian measures:

a. Compensating variation – Throughout the change in provision, this is the money income necessary to keep the individual at his/her initial level of utility. If the individual stands to gain from the provision then the compensating variation will be the amount of money that they are WTP to ensure the gain occurs. If the individual suffers a welfare loss then the compensating variation will be the amount the individual is WTA in order to tolerate the loss.

b. Equivalent variation – this is the amount of money needed to maintain an individual at his/her post change utility level. For a proposed gain the equivalent measure shows how much an individual is WTA to forgo the gain. For a welfare loss the equivalent variation measure shows how much an individual is WTP to prevent the loss.

Figure 2.2 illustrates the differences between a Marshallian and Hicksian demand curve.

Figure 2.2 Marshallian and Hicksian Demand Curve

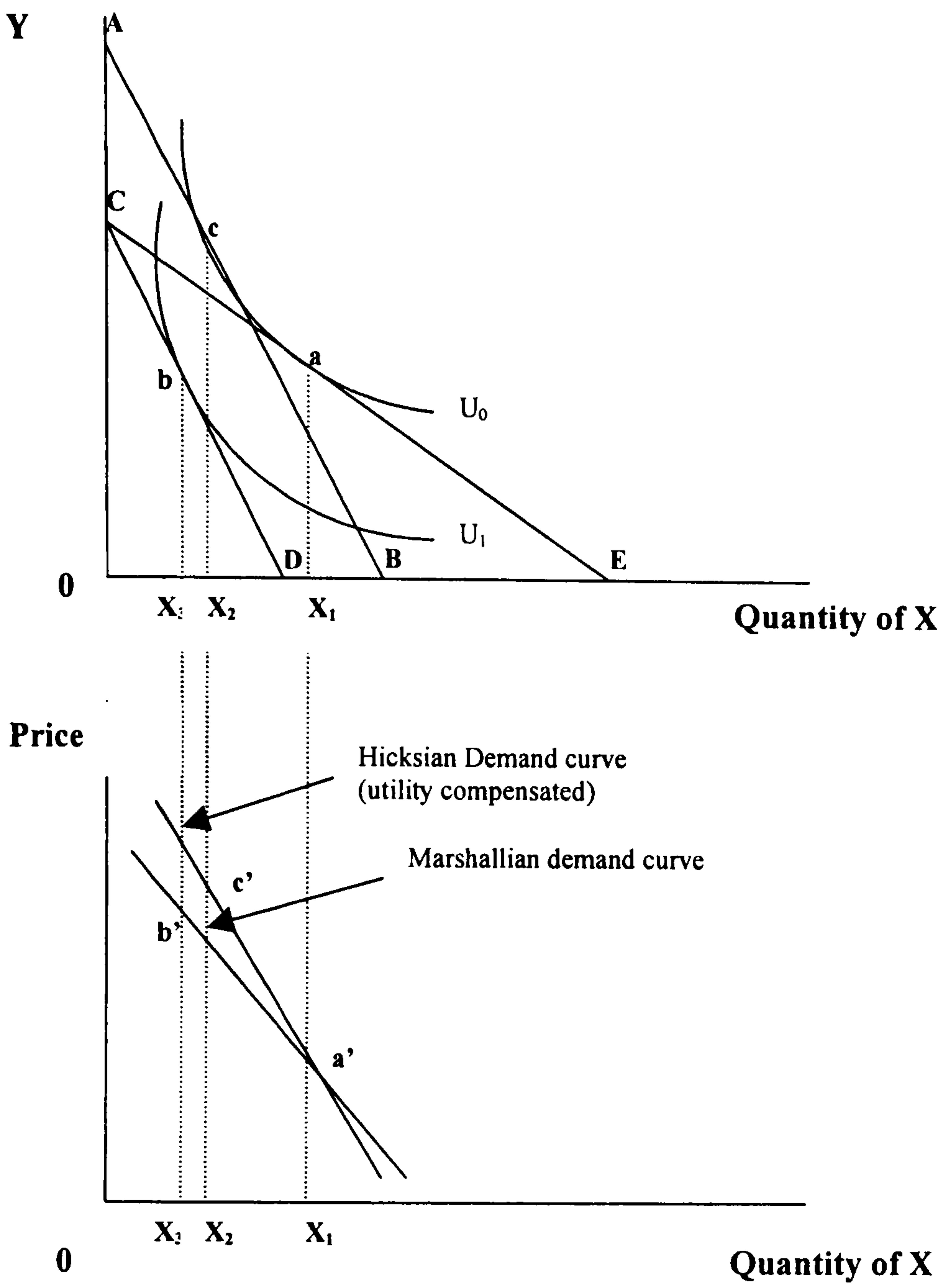


Figure 2.2 (a)

Figure 2.2 (b)

The maximum amount of money an individual would be WTA to tolerate a price increase is the amount required to return him to the same level of utility he enjoyed prior to the change in price, *the compensating variation.* Figure 2.2(a) presents the budget constraints and the indifference curves (level of utility), for a consumer who faces a world where there are only two goods X and Y. The budget constraint AB represents a higher income level than the budget constraint CD. The slope of the consumer's budget constraint is negative and depends on the price of good X relative to the price of good Y. If the price of good X rises, the slope of the budget constraint will become steeper from budget constraint CE to budget constraint CD.

If the consumer originally faces budget constraint CE, this means that they are at point a on the indifference curve U_0. If the price of good X doubles, the budget constraint becomes steeper to CD, the consumer moves to a lower indifference curve to point b (U_1). The increase in price of good X results in the consumption falling from X_1 to X_3. If the consumer were willing to accept an amount of money sufficient enough to compensate them for the loss in utility, this would shift the budget constraint in a parallel movement from CD to AB. The consumer would now choose point c on the original indifference curve (U_0) rather than point a and would now consume X_2 amount of good X. The compensating variation is the amount of money the consumer would be WTA to tolerate the price increase in good X (to return him to his original level of utility, U_0).

The total effect of the price increase on the consumer demand can be decomposed into two separate effects: a compensated substitution effect and an income effect. The compensated substitution effect is represented in Fig 2.3a as the change in demand from X_1 to X_2. This is the effect of a change in the price of X on the demand for X if the individual were exactly compensated for the loss of utility. The income effect is represented by the change in demand from X_2 to X_3 and results because the increase in the price of good X reduces the consumer's disposable income. If the price of good X rises, assuming it is a normal good, then the consumer's disposable income will fall and the consumer will purchase less of it. Hence, X_3 is smaller than X_2. The income effect will cause the consumer to reduce his demand for the good.

The line in Fig 2.2 (b) that connects points a' and b' is a conventional demand curve known as the Marshallian demand curve. The Marshallian demand curve incorporates both the income and the substitution effects associated with price changes in good X. The demand curve that connects both points a' and c' keeps utility constant throughout the price change in good X, it only incorporates the substitution effect associated with price changes. This demand curve is known as the utility compensated or the Hicksian demand curve. Hicksian demand curves are usually more steeply sloped than Marshallian demand curves.

Consumer surplus can therefore be measured using either the Hicksian demand curve (income compensated) or the Marshallian demand curve. The Hicksian consumer surplus can be thought of as the Marshallian consumer surplus measure calculated from demand curves where utility is held constant. According to Willig, if we adopt the simple measure of surplus as the area under the Marshallian demand curve then the difference in consumer surplus (compared to the Hicksian measure) will not be particularly large (Willig 1976). However if it is the value of the proposed change that we are attempting to measure, then there is a reasonable consensus that the measure we should be using is the area under the Hicksian demand curve as this compensates for the income effect (Pearce 1983; Donaldson 1995).

2.6 Health care demand curves

Measuring the costs and benefits of a programme is fairly straightforward when the demand and supply curves are known. However when they are not known, the valuation becomes more complex as the analysts have to find alternative ways of measuring benefits. These practical problems of measuring benefits are caused in the

UK health care market as health care services are offered at zero price. A conventional demand curve therefore is prevented from forming as individuals in society are all consuming at zero price. There are some sections of the health economy where individuals can choose to go private, e.g. private screening, and in these circumstances a demand curve can be estimated however in the majority of cases, health care is consumed 'free at the point of use'. Where a demand curve is prevented from being defined, economists have to look towards using alternative measures that capture the demand of individuals. Given this scenario, there are two main approaches to measure the demand for a programme, the first is through the revealed preference approach, the second, the stated preference approach, each will now be discussed in turn.

2.7 Revealed Preference Approach

The revealed preference approach, sometimes referred to as the indirect approach, uses practical methods to measure the benefits of a programme that are based on observing individual behaviour. Two revealed preference methods are discussed:

1. The hedonic pricing method.
2. The travel cost method.

2.7.1 Hedonic Pricing Method

The hedonic pricing method assigns values to the attributes of a given product. By adding or removing the attribute the affect on the consumer surplus can then be estimated. To explain this more clearly, a 'well-woman' clinic can be used as an example. 'Well-woman' clinics have many features e.g. size, location, friendliness of staff, waiting time, ambience etc. If we assume a scenario where patients have to pay to attend this clinic, the price they are willing to pay will be a function of all these attributes:

Price to attend clinic = *f*(size, location, staff attitude, waiting time, ambience, etc.)

This function is called the hedonic price function and the hedonic price is the change in the total willingness to pay from a unit change in a given attribute, holding all other attributes constant. For example, the slope of the hedonic price function with respect to the location of the clinic indicates the relationship between the cost of attendance and the location of the clinic. The method therefore is trying to measure the change in consumer surplus for a given change in the location of the clinic. The change then in an individual consumer surplus can be aggregated across all individuals to obtain the total change in consumer surplus from changing the location of the 'well-woman' clinic.

However a disadvantage of this method is that individuals may value products for reasons other than the obvious attributes in the estimated function. An example of one such attribute is the 'non-use value' and the 'existence value' component (Hannemann 1994). Individuals may gain utility from health care programmes for various reasons other than their expected personal use. It is feasible that individuals will experience a welfare increase from knowing that a 'well-woman' clinic has had improvements, should you wish to attend (Kruitilla and Fisher 1975). The hedonic pricing method fails to capture this type of utility.

2.7.2 Travel Cost Method

The travel cost method (TCM) was developed principally in papers by Clawson (Clawson 1959) and Clawson and Knetsch (Clawson and Knetsch 1966). The method estimates the demand for a good by measuring the travel costs incurred to 'visit' that good. For example, the costs of visiting a screening clinic are more than just the experience of being screened and the time spent at the clinic, patients (and perhaps a companion) pay to travel to the screening clinic, this takes time and time that could have been spent elsewhere (opportunity cost of time). As each person visiting the clinic will travel a different distance, the total travel cost per person will vary. The usage of the clinic will also vary, this coupled with the travel cost provides an opportunity to make inferences about the demand curve for the screening clinic. The method involves the analyst drawing 'circles' around the site and the area between these circles are referred to as zones. Through analysing the travel information data that is usually obtained through the use of survey techniques, the analyst will assign a monetary value to each mile travelled. The consumer surplus for each zone is then calculated by measuring the area under the demand curve, and above the 'travel cost'.

The TCM has many disadvantages associated with the technique. One would not expect visits by an individual to depend only on travel and costs, income for example would vary across individuals and affect compliance rates. Also, it is feasible that time spent travelling to the clinic could have positive utility, and, if this is so, should be assigned a negative cost. People may use the screening appointment as an opportunity to do other activities, e.g. if the screening clinic was located in town, and the individual had the day off work, they may visit the clinic in the morning and go shopping in the afternoon - this poses a problem of how the travel cost for the day should be estimated.

An alternative to measuring the costs and consequences of a programme or a product change is to use a stated preference method.

2.8 Stated Preference Approach

The contingent valuation (CV) approach (direct approach) is a stated preference method designed to directly estimate welfare gains/losses as appropriate. How it works is that individuals are asked to consider a hypothetical scenario where they are asked to imagine that a market exists for the benefits or losses of a public programme. The exercise proceeds on the hypothetical contingency that such a market exists. Various design instruments can then be applied to ask individuals to state their willingness to pay (WTP) to ensure a welfare gain occurs or willingness to accept (WTA) to tolerate the welfare loss from the programme. The WTP or the WTA amount is then taken as a measure of the individual perceived value of the programme (i.e. the demand) which is then aggregated across all individuals. If individuals state a high (low) WTP amount then it is inferred that the demand for that programme is high (low).

2.9 History of CV

The use of the contingent valuation (CV) method to assign values to benefits for goods without clearly defined demand curves has been traced back to 1958 (Hanemann 1992). During the 1960s the application of the method was infrequent and almost always applied in the United States. It was not until the 1970s that the usage of the method had reached a point where it was recognised by the US Water Resources Council (1979) as a recommended valuation technique along with the travel cost method. In the mid 1970s the U.S. Environmental Protection Agency funded a program of research with the methodological aim of determining the strengths and weaknesses of the CV method. The 1980s experienced a dramatic increase in the popularity of the method, not only within the United States but worldwide. By 1990, the CV method was the most dominant technique for the valuation of non-market environmental costs and benefits.

The increase in the application of the technique brought about an intensification of debate regarding the reliability of the method. Design issues such as the elicitation method, the choice of framing for the question, the payment vehicle and the influence of a possible embedding or an anchoring effect were discussed. As a consequence the National Oceanic Atmospheric Association (NOAA) commissioned a report that evaluated the method. The review panel comprised psychologists and economists and included the Nobel laureate Kenneth Arrow. After careful consideration of the CV method, this report was produced in 1993 claiming that so long as studies adhere to strict guidelines (laid out in the report) then the method as a tool to measuring environmental benefits is satisfactory. These recommendations have been used to debate methodological problems of using the CV method in the health care field and will be discussed in depth in chapter four of the thesis.

Within health care, one of the first reported attempts to use the WTP method was a study that estimated the demand for increased car protection. This study found that 1,017 car buyers were willing to pay an average of $12 to their monthly car payments, if the added cost would save 6,000 lives a year (Robertson 1977). This study was published in the Journal of Community Health. From 1985 onwards there has been a steady growth in the number of published papers using the CV method within health care (Olsen and Smith 2001). The majority of these studies have been reported in clinical and public health journals and much of the CV work that has taken place over the last 20 years within the health care field will be discussed extensively throughout this thesis.

2.10 Conclusion

It is clear that the CV method is well established within the field of environmental economics however it is only relatively recently that it has become popular within health economics. Its application within health economics has been the subject of much debate with regard to the methodological components. Arguments for using the CV method instead of alternative outcome measures within health care relate to the fact that it has theoretical underpinnings within welfare economics; the ability for the method to encompass a much wider measure of benefit; and the practical issue of measuring costs and outcomes within the same unit.

Chapter three expands these arguments and explains why a health analyst may choose to adopt the CV measure when evaluating health care programs. In particular, aspects of health care that are encompassed by the CV measurement are talked about with an emphasis on how the available alternative measures of outcome fail to capture these features.

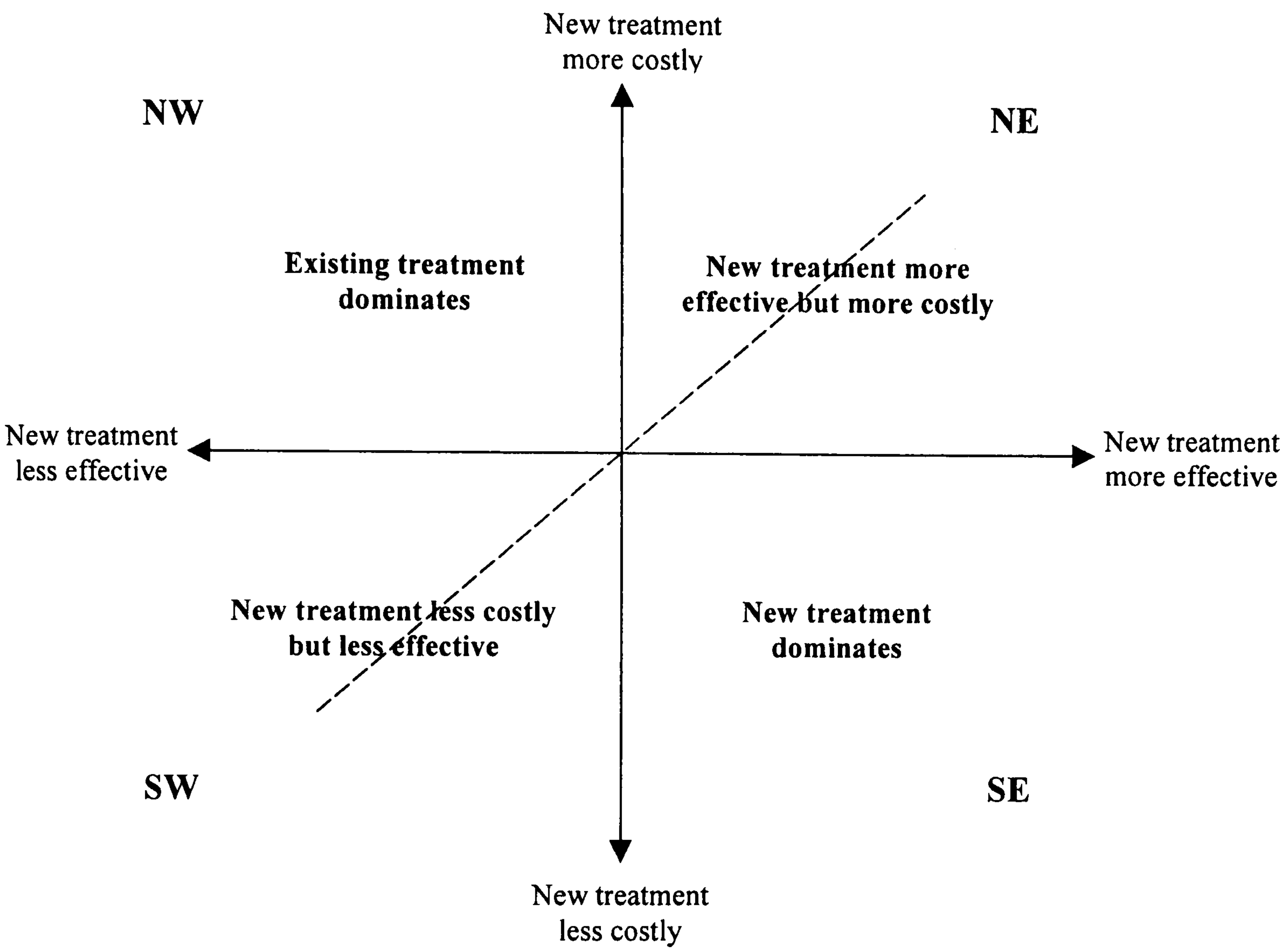
Chapter three – Use of WTP in Health Care

3.1 Introduction

Interest in economic evaluation as a tool to aid health care decision making has grown steadily over recent years. This interest has grown due to an increasing recognition that publicly funded health care systems have limited resources making it essential to consider the opportunity cost of decisions made. Initiatives such as the development of the National Institute for Clinical Excellence (NICE) in the United Kingdom (UK) provides evidence that the Government are considering it more important to make decisions within an economic context. NICE has been set up with an objective to provide guidance to the UK National Health Service (NHS) on the clinical effectiveness and the cost-effectiveness of health care technologies.

The purpose of an economic evaluation is to provide advice to decision makers on health care interventions that provide good "value for money" and therefore assist in the allocation of health care resources. Within the decision making framework an economic evaluation is set up to identify the costs and consequences of alternative health strategies. When considering the cost-effectiveness of a new health care technology, it is very rare that the technology will replace all existing and established treatments. The economic evaluation therefore will consider the cost per unit of benefit from moving patients from one treatment (usually the existing treatment) to a new treatment. The results will thus represent the additional cost from providing extra units of outcome. This is best explained by means of a diagram illustrated in figure 3.1.

Figure 3.1 The cost-effectiveness plane



The cost-effectiveness plane represents graphically the relevant dimensions of an economic evaluation. When comparing health care treatments the difference in effect is plotted along the horizontal axis, and the difference in cost on the vertical axis. Each of the quadrants are usually labeled with compass bearings. In the SE and NW quadrants, one intervention is clearly more effective and less costly than the comparator therefore the decision is straightforward. In the SW and NE quadrants the decision involves a trade-off concerning the additional cost of additional effectiveness, and the maximum acceptable level of cost effectiveness becomes more important. Incremental cost-effectiveness ratios represent the additional costs that one service or program will incur against the additional benefits that it delivers. An example of incremental analysis is represented in table 3.1.

Table 3.1 Incremental Analysis

Option	Total cost	Total outcome	Average cost-effectiveness ratio (C:E)	Incremental cost-effectiveness ratio (ΔC:ΔE)
Option 1	C_1	E_1	C_1: E_1	
Option 2	C_2	E_2	C_2 :E_2	C_2-C_1 :E_2-E_1

Average cost-effectiveness ratios are calculated by dividing the average cost of treating patients by the average outcome (C_1/E_1 or C_2/E_2), whereas incremental cost-effectiveness ratios are obtained by dividing the incremental cost by the incremental benefit ($(C_2-C_1) / (E_2-E_1)$).

There are different techniques available to health economists to assess the costs and benefits of alternative health care interventions. During the 1960s and the 1970s, cost of illness studies dominated (Johannesson and Jonsson 1991) however this method is limited as only the costs incurred by society are considered, no attention is directed towards the effects. As shown in the cost-effectiveness plane, when operating in a world with scarce resources it is important to consider both the costs and effects of an intervention, and to do this, there are three types of economic evaluation. These are cost effectiveness analysis (CEA), cost utility analysis (CUA) and finally, cost benefit analysis (CBA).

3.2 Cost-effectiveness analysis

A cost-effectiveness analysis (CEA) method is designed to measure the incremental benefits (Ei) in 'natural' units that are then compared with the incremental costs (Ci). Examples of the effectiveness measures are 'life years gained' (the most frequently used), number of 'successfully treated patients', 'cancers detected' etc. Health care programmes are ranked according to their incremental cost-effectiveness ratio [Ci/Ei] and selected from the lowest to the highest ratio until the budget is exhausted. A CEA is used to determine if the cost of health improvements is acceptable in terms of the cost of alternative health improvements in other disease areas. To allow comparison between different health care interventions, the effects have to be measured in the same unit.

Although CEA is a widely used method for economic evaluation (Johannesson and Meltzer 1998), there are suggestions that it has the following limitations:

1) The concept of risk is not incorporated into the cost effectiveness measurement. The CEA does not seek to measure individuals' attitudes towards different levels of risk associated with health care programmes. For example, with a screening programme, groups of individuals will have different levels of perceived own risk to that disease which may affect the final outcome of the programme. It is the perception of different attitudes to risk that CEA is limited in measuring.
2) There are problems in identifying which costs to include, i.e. medical/non-medical or productivity costs. This varies from study to study. There is a debate as to whether future health care costs brought about by prevention and extra costs (that are not related to the treatment) but due to increase in life expectancy, should be included in the analysis. The inclusion of which type of cost usually depends on

the perspective taken for the evaluation. If the study takes on a societal perspective then costs relating to morbidity and side effects should be included however if a decision-maker perspective is taken then there is an argument for the exclusion of these costs. As there is no agreement about what costs should be covered it makes it difficult to compare cost-effectiveness ratios from different studies.

3) A CEA does not guarantee comparability across studies with different dimensions. For example, the health care analyst cannot compare outcomes associated with treatment of myocardial infarction (where effects are expressed in 'life years gained') with that of skin rashes (where effects are expressed as 'reduction in discomfort'). There are certain disease areas where life expectancy is not a useful measure of the benefits of a treatment and the impact on the quality of life is more important. One example is a hip replacement operation, this intervention is unlikely to have an effect on the patient's life expectancy but it will have a considerable impact on their quality of life. When using a CEA to measure the costs and effects of a hip replacement therefore the unit of measurement will be the improvement in the mobility of the hip, making it impossible to compare a hip replacement program with another program that measures the benefits in 'life years gained'.

4) A CEA struggles to cope with health care interventions that produce gains in more than one dimension. For example, improvement in an individual's quality of life or general well-being is not captured by 'life years gained'. The costs can be divided up into the various areas where benefits are produced but this is always going to be arbitrary and the results are difficult to interpret.

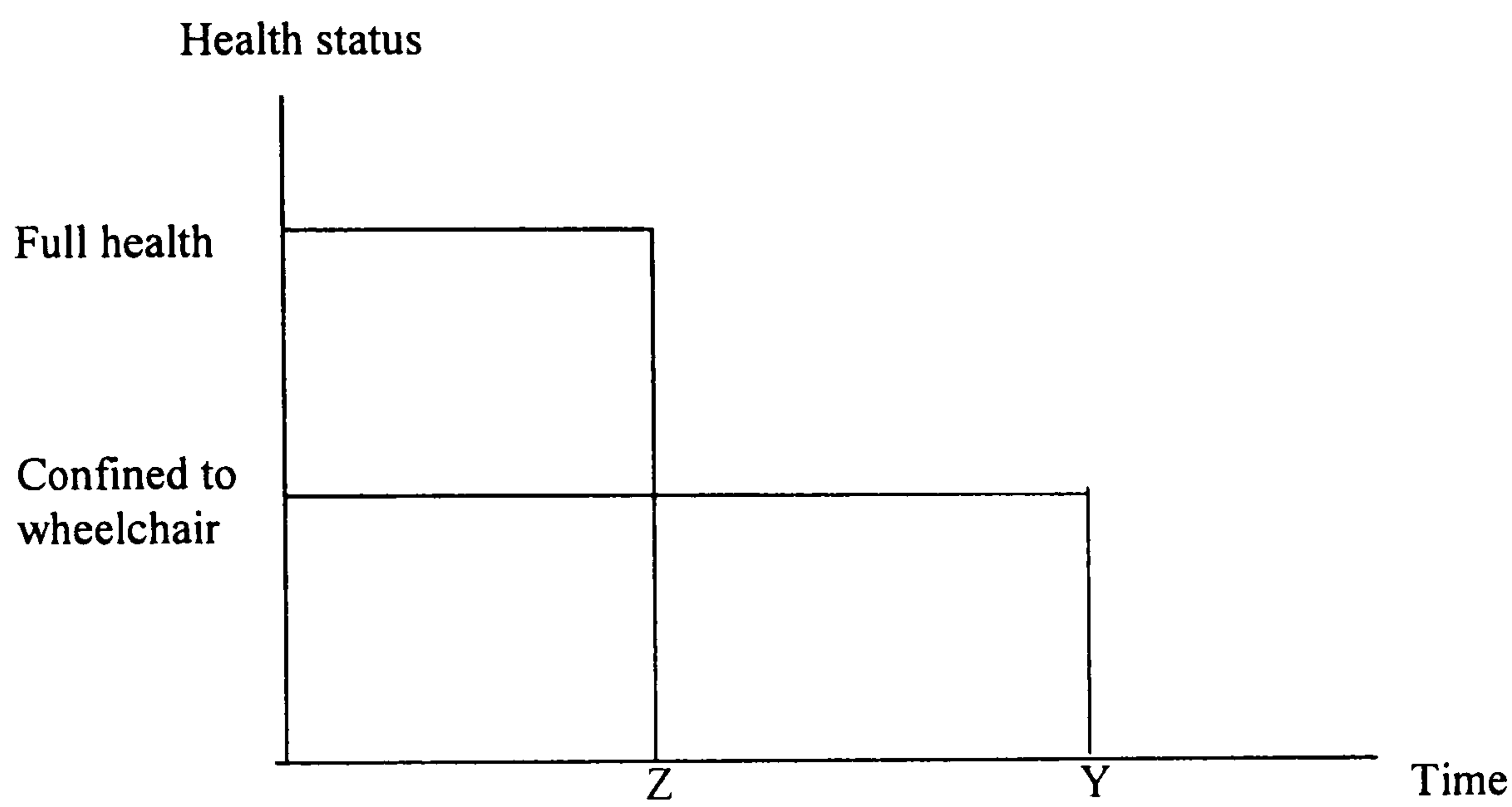
3.3 Cost-Utility Analysis

As discussed, a CEA measures the effects of a health care intervention in physical units an extension of this measure would take into account a person's quality of life (QOL). A cost-utility analysis (CUA) compares the incremental cost of a programme with the incremental health improvement where this health improvement is measured in quality-adjusted life years (QALYs). The quality adjustment is made based on the utility weight that reflects the desirability of a given health state. If a health care intervention affects the morbidity of patients then a CUA is the appropriate form of measurement to adopt as it takes account of this morbidity. The optimisation procedure is similar to a CEA only with a CUA, the aim is to achieve the maximum number of QALYs for a given budget. The QALY is obtained by weighting the time spent in a certain health state with a specific utility value which gives us the QALY weight on a scale of 0 (death) to 1 (full health). These utility weights are obtained by asking individuals about their preferences for the states. There are three separate methods used to obtain utility values, the rating scale (RS), the time trade off (TTO), and the standard gamble method (SG).

The most common version of the RS technique is the visual analogue that provides the patient with a scale with 0 indicating death and 100 representing full health. If the patient rates a health state as being at 70 then the QALY weight is calculated as 0.7 (i.e., 70/100).

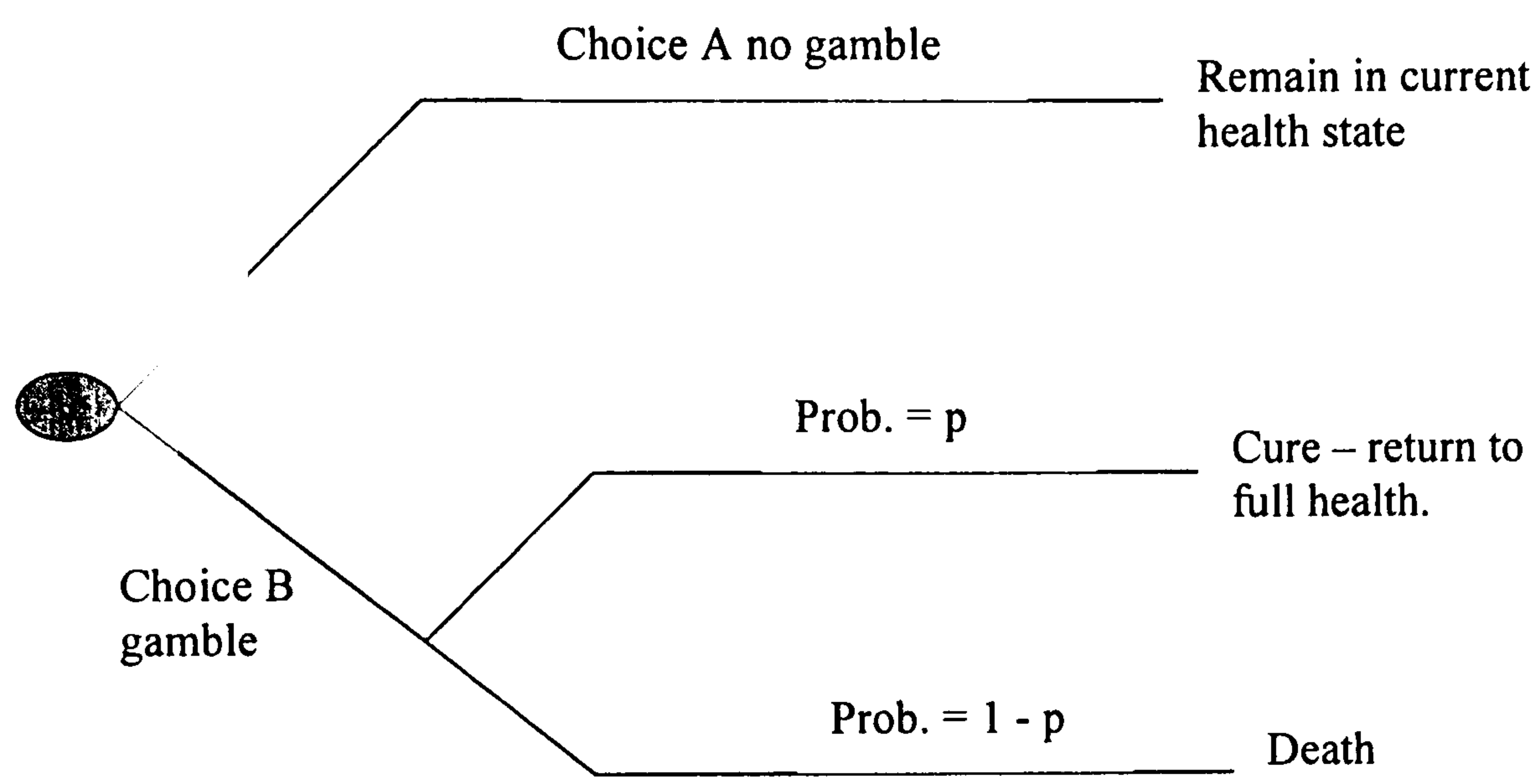
The TTO method provides the respondent with two alternatives. The first is to remain in the current health state for Y years, the second is to return to full health for Z years where Z < Y. The time Y is varied until the respondent feels that they are indifferent between the two alternatives. An example of this technique would be where the respondent values 2 years being confined to a wheelchair as equivalent to 1 year in full health, the QALY weight attached to being in a wheelchair is therefore 0.5 QALY (i.e., 1/2) over the course of one year. The TTO method is illustrated in figure 3.2.

Figure 3.2 Illustration of the time trade off method



The SG method for obtaining QALY weights is regarded as the 'gold standard'. Again the respondent is given a choice between two alternatives, the first is to remain in a current health state (intermediate between wellness and death) for a certain number of years. The second involves an element of risk, the respondent is told that they can receive an intervention (with an associated risk) with a probability of *p* that it will be successful and a probability of 1-*p* that it will fail and the respondent will die immediately. The probability is varied until the respondent is indifferent between the two alternatives, i.e. indifferent between the risky intervention and the certain number of years in the current health state. The QALY weight is then calculated by the level of *p*, e.g., if the respondent is indifferent between a certain number of health years in the current health state and the risky intervention with a probability of success of 0.6, the QALY weight attached to the health state is 0.6. Figure 3.3 represents this diagrammatically:

Figure 3.3 Illustration of the standard gamble (SG)



When using the SG method, visual aids and props are normally required to explain the method. Respondents however find the SG difficult to understand as they usually struggle with the probability concept.

An inherent advantage of using a cost-utility analysis to evaluate health care interventions is the common unit of effectiveness, the QALY, making it possible to make comparisons across health care programs that treat different classes of disease. The CUA however is not without its own set of limitations:

1) Given that there are different methods that can be used to calculate the utility levels (the RS, TTO and SG) this gives the potential for variations in the

estimation of the QALYs across studies. Where studies have used different methods it is difficult to make comparisons between them for the allocation of resources. Nord (1992) reviewed ten articles between the period 1976 and 1991 that compared utilities for different health states measured by different methods - on average, the SG and the TTO methods give higher values than rating scales and the SG method gives higher values than the TTO method (Nord 1992). To allow comparisons across different studies, there needs to be a consensus regarding the most appropriate QALY estimation method to use.

2) There are also difficulties over the interpretation of QALYs for individuals with differing socio-economic backgrounds. In particular, a QALY has been criticised as being 'ageist' as it favours health care programs that improve health for the young over the old (the young have longer to live therefore more QALYs to gain). Also, it is not agreed in the literature that there is a constant relationship between remaining years of life and improvement in quality of life.

3) Levels of income have a systematic effect on the rate of life expectancy (Robine and Ritchie 1991; Evans, Barer et al. 1994; Wilkinson 1996). This means that the underlying distribution of income within a population may affect the QALY calculations. For analytical purposes it is possible to assume the same life expectancy for all individuals within the population, but this departs from the notion of opportunity cost if the assumed life expectancy differs from the actuarial life expectancy for individuals in the population. The opportunity cost of losing 1 QALY will be greater for an individual with 10 years to live than for an individual with 50 years to live, assuming a constant life expectancy for every individual ignores this opportunity cost.

The calculation of life expectancy plays a large role in QALY calculations, and given that life expectancy is related to income it is logical to assume that individuals in poorer income groups have less time to trade for quality of life improvements than high income groups. This means therefore that the marginal valuation of life years in a given health state can be systematically different between different income groups.

Traditionally, the economic belief is that health care has no value in use (Donaldson and Gerard 1993), put differently, it is assumed that no one particularly enjoys receiving health care and that they only demand it for the benefits derived from it, i.e. health improvement. Put more accurately we could say that no one particularly enjoys the consumption aspects of health care but they may enjoy the investment benefits that lead to an increase in health status. However, the benefits of health care can go beyond health gain making measures such as 'life years saved' and 'QALYs' too narrowly defined to incorporate <u>all</u> benefits from a health care intervention. According to Mooney, there can be many other attributes of health care that enter an individual's utility function (Mooney 1994). Later sections of this chapter will explain the aspects of health care that may directly affect an individual's utility level but are not captured by the outcome units used in CEA and CUA.

3.4 Cost-benefit analysis

A cost-benefit analysis expresses the effects of a health care intervention in monetary terms. For many years the most commonly applied method of benefit measurement within a CBA was the 'human capital approach' (Griggs and Mushlin 1998). The method works by equating the value of life extension and losses due to morbidity with foregone earnings that are discounted to their present value. Therefore the loss to the economy would be expressed as:

$$L = \sum Y_t P^t_r (1+r)^{-(t-r)}$$

Where Y_t is equal to the expected gross earnings by the person in the t^{th} year and the P^t_r is the probability in the current year (r^{th} year) of the person being alive in the t^{th} year. Applying this method to health care means that the value of a person is defined by their potential, inherent net product and this net product is evaluated using the income gained through employment.

A slight departure from estimating the lost productivity due to illness by earning capabilities is to use the insurance principle and base the estimate on the premium an individual is willing to pay to insure themselves against dying as a result of a specific activity. This method assumes a straight-line relationship between the probability of a person being killed and the sum that he would pay to cover the risk. The individual will pay a premium equal to *y* corresponding to the additional risk *p*, hence the value he places on his life is estimated as *y*/p. For example, if the probability of being killed in air travel were to be reduced from the existing figure of 0.0000017 per trip of 500 miles to zero, a person who values his life at $400,000 should be willing to pay sixty-eight cents to reduce the existing risk to zero.

As is fairly obvious when evaluating health care, the 'human capital approach' has some limitations. These are:

1) It is suggested that the method excludes 'costs' of pain and suffering. The question then to ask is whether these 'costs' of pain and suffering are actually impacting upon resource use, i.e. use of pain relievers, physiotherapy appointments. If so, then the costs associated with pain and suffering should be allocated to the cost-side of the cost-benefit ratio. If however the 'costs' of pain and suffering refer to the psychological burden of experiencing pain or being unable to work then this will be allocated to the benefit-side of the cost-benefit ratio as it is having no impact on resource use. This is an important point, as the allocation of these 'costs' will have a big impact on the final cost-benefit ratio. Where the costs are placed on the benefit-side of the ratio, occasionally the wage calculation may be supplemented by further calculations that are made to measure additional aspects such as the suffering of the victim, the loss of his utility or the bereavement of the family (Kneese 1966; Ridker 1967). However this is very rarely done and according to Mishan (1971), it may be more a response to an uneasy conscience about the methods used in the human capital approach.

2) Since the technique uses an equation estimated by individuals' wages, the method assigns no cost of illness to persons who are retired or live off non-labour income. This means that within the health care industry, health measures that will be applied to those in employment lead to a better cost benefit ratio than when they are applied to the unemployed.

3) There are earning differences across the population, but this difference may reflect wage discrimination instead of variation in productivity. Sometimes the payment for work does not reflect the contribution of the services carried out by the

individual. There exist imperfections in the market that lead to a number of situations where different values are placed on the performance on individuals, for example, the differences in male and female pay for the same level of qualifications and experience. Based on this premise, it can be argued that the human capital method is biased as it favours white, adult males who are felt to have the highest earning potential.

4) The precise social rate of discounting is unknown. According to theory, the appropriate rate of discounting when making social decisions would use the social rate of time preference, that is, the rate at which society is willing to forego consumption today for greater consumption tomorrow. However the individuals that make up society will each have a different discount rate that is likely to differ from the social discount rate. This is because individuals play a dual role in society, firstly, that of consumers who wish to maximise their own utility, and secondly, that of citizen that has some responsibility for the society that they are a member of. It is unlikely that the individual's preferences for resource allocation will be the same from both perspectives.

5) Cost savings from averted future illness are not subtracted from medical care outlays. It is unclear how far into the future health care costs should be considered, if the intervention increases the life expectancy of the individual, hence future earning capacity then the future costs of illness/prevented illness may need to be incorporated into the analysis.

6) With the insurance principle, the implied assumption of linearity between the probability of a person being killed and the sum that he would pay to cover the risk is implausible. This effectively assumes that with certain probabilities an individual will agree to go towards certain death (if an individual takes out a £100,000 insurance

premium offering a four to one survival this implies that the individual will accept £500,000 for certain death). However it is plausible that in fact the relationship is linear but only up until a limit. Once the risk of death exceeds a certain level the relationship will become discontinuous hence no amount could compensate an individual for certain death.

7) The trouble is that even if the linear assumption for the insurance principle was plausible, insurance policies in effect make provision for <u>others</u> in the event of a death, not the individual concerned. The premium amount therefore will depend ultimately on the concern of the individual for his family and dependents, not on the value he places on his life.

In fact, the 'human capital approach' has been widely criticised because of its inconsistency 'with the basic rational of the economic calculus used in cost benefit analysis' (Mishan 1971) which is the aim of enhancing welfare based on the preferences of individuals. Johannesson argues that the approach has no base in economic welfare theory and by using it the intrinsic value of life is ignored (Johannesson and Jonsson 1991). To take the view that health only has value in adding to 'national output' is inconsistent with welfare economics as welfare economics takes account of the value of leisure time and in other activities that are not measured in the gross national product (Pauly 1996).

Due to these limitations, measuring benefits in monetary terms using the human capital approach has become rare and consequently there has been a movement towards refining alternative techniques for measuring outcomes in a CBA such as the CV method.

3.5 Why use the CV method in health care?

Chapter two introduced the CV method, as described, it is a 'survey-based approach for eliciting consumer's monetary valuations for programme benefits for use in CBA' (Diener, O'Brien et al. 1998). The method seeks to elicit by how much people collectively value goods. The method works by asking individuals to state either their willingness to pay (WTP)(where there is a welfare gain) or their willingness to accept (WTA)(where there is a welfare loss) for the programme to go ahead. The most common application of the CV method for the evaluation of health care studies is to ask a sample of individuals for their WTP value although the measurement of WTA has been occasionally applied (Diener, O'Brien et al. 1998). The terms WTP and WTA can be used interchangeably to refer to the CV method.

As a technique of measuring health care benefits, CV has gained popularity in the last two decades (Diener, O'Brien et al. 1998). The CV method has been advocated because of its theoretical relationship with welfare economics (Olsen 2000), however the main advantage of the method is that it encompasses a wider measurement of benefits than the alternative QALY or 'life year' saved units that are used in CUA and CEA. Willingness to pay is a measure of '*strength of preference*' for one health care intervention/program over another. Asking people to indicate which health care intervention/program they prefer reveals the '*direction of preference*', however, if they are then asked to state their maximum WTP for the health care interventions/programs this provides the individuals with the opportunity to reveal their '*intensity of preference*'. By summing up the WTP values across the whole population, the health care analyst can then draw conclusions about how much one option (program) is valued in relation to another.

The CV method has advantages over the alternative methods of measuring benefits in health care (the standard clinical outcomes used in CEA and QALYs used in CUA). The main benefit of adopting the CV method is that it provides the scope to measure aspects of health care beyond that which the CEA and CUA outcome techniques measure. These aspects include:

- the demand for information
- process utility
- option value
- risk assessment

The remaining part of this chapter will discuss each of these aspects in turn. During these discussions examples of empirical research will be used to clearly demonstrate that these aspects are important and are commonly left out of the CEA and CUA ratios. Therefore, the discussion will demonstrate the overall strengths of choosing a CV method to measure the benefits of health care.

3.6 Demand for information

Some health care interventions may be derived from the demand for information (Ryan 1996). This relates to the decisional aspects of health care. Like most health care systems, the NHS is characterised by an asymmetry of information between the patient and the doctor. This means that the doctor has the knowledge to diagnose what is wrong with the patient and make the decision of which treatment best meets their needs and the aim is that the doctor should demand what the patient would have demanded, had they had the same knowledge as the doctor. The patient looks to the doctor to demand health care on behalf of them. Where the health intervention provides information to the patient, the patient can gain utility from being better informed and becoming part of the decision-making process. The patient translates the information received into knowledge and ultimately feels better from gaining knowledge from the health care service. This has been demonstrated by a study conducted by Berwick and Weinstein (Berwick and Weinstein 1985) that investigated the demand for various amounts of information that is generated from an ultrasound scan. The types of information included were health of the foetus; the health of the mother; the expected delivery date; multiple pregnancy; the sex of the foetus; a visual image of the moving foetus on the screen; and a hard copy of the image. Results of the study indicated that respondents were willing to pay ten times more for this information over the local charge for a traditional ultrasound scan (where less information is provided). The study showed that the respondents valued the type of information used to make decisions and that utility can be gained from the information provided.

Another study looked at the value patients' place on negative cervical smear test results. Traditionally negative (normal) results from a clinical activity are assumed to have no value (Grimes 1988). However patients gain utility from being certain, therefore one can view the demand for information from a clinical activity as being a derived demand from the demand for certainty. This is demonstrated by a study that explored the value of a normal result of a cervical smear test to a woman and it found that out of eighty-four responses, after exclusions were made, sixty-nine women were prepared to pay a mean value of £8.25 for a negative smear. It was clear therefore that the patients valued the information gained from a screening test, even if the result was negative.

3.7 Process utility

Many studies to date have attempted to estimate the 'use value' in health care, (Williams 1985; Johannesson 1992) employing measures such as QALY's or WTP. It is fair to assume that no one actually enjoys receiving health care therefore health care has no 'positive' value in use. However, from the perspective of the individual, the process of health care may actually entail negative value in use. This means that individuals, depending on their preferences, can gain different levels of utility from different types of care. For example, some patients prefer less invasive treatment over a longer period of recovery time to procedures with a quicker recovery time but a more invasive nature or individuals may prefer home-based to hospital care. In other words, the process of care can have a direct impact on the level of individual utility. This is important when comparisons are being made between two health care programmes that have equivalent outcomes, but different processes of care. Potential patients may have strong preferences over which process of care they prefer. Patient satisfaction forms an integral part of many a non-economic evaluation of medical care services (Hall and Dornan 1988). There have been numerous studies of patient satisfaction (Cleary and McNeil 1988), and it is widely agreed that the data should be used to focus and facilitate quality improvement efforts (Cleary 1999). A good example is screening for disease, as this may entail psychological costs. It is common for individuals to experience negative feelings when informed that they have a positive result in a screening test. Not all the positive screening results however will lead to a positive diagnosis (false positive result) and some individuals, after further testing, will later discover that they do not have the disease. However research has shown that once the seeds of doubt are sown it seems difficult to remove them (Marteau 1989).

Patient satisfaction is rarely incorporated into an economic evaluation. If individuals are asked for their willingness to pay for alternative programmes then the utility gained (or lost) from the process of care can be included in the measurement. The CV approach allows individuals to take account of all factors which are important to them in the provision of the service (Ryan 1996).

Donaldson and Shackley applied the willingness to pay technique to investigate if process utility exists (Donaldson and Shackley 1997). The study sought to measure preferences for laparoscopic rather than conventional cholecystectomy. The study sample were split into two groups, one group received a description of the differences between the two treatments in terms of outcomes only, whilst the other group received information on differences in the process of treatment as well as on differences on outcome. It was hypothesised that information on the process of care would lead to a higher WTP. However the results of the study indicated that the opposite was the case. What actually happened was that the respondents that received the information on process of care gave a lower WTP for the laparascopic surgery. Several reasons are stated why this may have occurred. Individuals may not want to know what they will have to go through as part of an operation therefore the descriptions actually affected them in the opposite way to a priori expectations. The study also sheds light on the difficulty of separating out descriptions of process and outcome. If process is anything which takes place during surgery, and recovery, then the outcome descriptions used in this study covered aspects of process, i.e. faster recovery, quicker return to work, shorter hospital stay and less pain. If outcome is defined as the state the patient is left in after the surgery, then the only outcome definition for laparoscopic cholecystectomy is a smaller scar. Therefore the authors suggest that the 'process and outcome' description actually comprised six aspects of process and one of outcome, and it could be argued that they did not test the hypothesis that they set out to. However the authors do conclude that process utility does exist but not in a form that could be established by this study.

Using the method of willingness to pay, Ryan (1996) established the importance of factors beyond the traditional medical benefit (i.e. conception) of assisted reproductive techniques (Ryan 1996). Process attributes defined as potentially important in this study comprised attitudes of staff; location of clinic; waiting time; continuity of contact with staff; speed of investigation; and costs of treatment. The results suggest that, even if couples leave the treatment childless, there is some value gained from going through the service. The study found that 83% of respondents agreed with the statement: *' One of the reasons I am trying (or tried) IVF is so that in later life I will know that I have tried everything possible to have a child.'*

Another study conducted by Donaldson et al. explored the strength of preference of women for a midwife-managed delivery unit *versus* care in a consultant-led labour ward using the WTP technique (Donaldson, Hundley et al. 1998). The total WTP by women who preferred a midwives unit was £10,030 compared to £6,070 for women who preferred labour ward care. Therefore, ceteris paribus, the women in the study have expressed a clear preference for the midwives unit through their WTP values. Reasons for the preference were *'more homely and relaxed', 'more personal' and 'continuity of care'*.

Gibb et al, in 1998, used a similar method to explore the strength of preference for different abortion methods; medical abortion *versus* surgical vacuum aspiration (Gibb, Donaldson et al. 1998). The medical abortion is a procedure that does not require a general anaesthetic and such a long stay in hospital. Earlier studies have shown there to be no difference between the two procedures in terms of safety, effectiveness, effect on future fertility and psychological sequelae. The majority of women (68%) indicated a preference for the medical method primarily due to the fact it did not require a general anaesthetic. However the WTP values given for both methods were largely similar. The authors conclude that although more women prefer the medical method, the intensity of preference, for the majority of women, is similar to those who favoured the surgical method. Therefore if the service was to be provided in a manner that satisfies the preferences of all women then the medical method should be made more widely available, but the surgical method should remain an option for those women who have a preference for it.

3.8 'Option' value

Individuals at risk of a disease may gain utility from knowing that health care exists for treatment in the future. This is known as 'option value'. The most distinguishing feature of the demand for health care is its unpredictable nature (Arrow 1971). People cannot predict the future therefore do not know when they are going to be faced with an illness and therefore require health care treatment. Knowing that it is available, should they need treatment, enhances their utility level. The CV method can take account of the option value as if an individual gains utility from knowing that a health care treatment exists then the maximum WTP for that treatment will reflect that.

To take account of the option value when evaluating a treatment the survey question can be designed to provide respondents with information about the probability of contracting the disease as well as a detailed description of the type of treatment that will be available. Neumann and Johanesson examined levels of WTP for the 'option value' by providing respondents firstly with a scenario in which they had a 100% chance of being infertile (therefore seeking a WTP estimate solely for outcomes associated with invitro-fertilisation (IVF) treatment) and secondly, by then varying the risk of being infertile to 10% and eliciting a WTP value for IVF treatment should they require it. The difference in the ranges of WTP values for each scenario then reveals the respondents perception of, should they become infertile, the availability of IVF treatment. The results showed a WTP range of $US 17,730 and $US 63,896 for when the respondent is 100% infertile and a range of $US 2,006 and $US 45, 865 for a 10% risk of being infertile. Logic would predict the WTP level to rise along with the corresponding risk of being infertile and it is clear from the WTP values that with a risk of 10%, respondents value the option of IVF treatment being available (Neumann and Johannesson 1994).

3.9 Risk reduction

The CV technique can also be applied to measure how individuals perceive their risk towards certain diseases:

'The measure incorporates the individual's preference for risk aversion, valuation of pain and suffering, preference of postponing death, and appreciation of reducing risks to life and health by small amounts (Muller and Reutzel 1984).'

Dickie et al used the method to measure individuals WTP to avoid skin cancer (Dickie and Gerking 1996). The study design used individuals' reservation prices for sun protection products to examine the link between risk beliefs and WTP to reduce risk. Through labels, the product was described to the individuals, it offered three types of protection for solar radiation (aging/wrinkling, suntanning/burning, risk of skin cancer). The study randomly assigned the three types of protection across the sample and obtained reservation prices for each level of protection. This allowed skin cancer risk reduction to be separated from the value of the other product characteristics. The study concludes that the willingness to pay for each unit of risk reduction is positive and increases with income.

Lee et al used a different approach to measure the impact of risk information on patients WTP for autologous blood donations (Lee, Liljas et al. 1998). The sample of patients were randomly assigned to one of two groups. The first group was informed of the risks associated with blood transfusions, the second group based their WTP on their own prior knowledge. By providing individuals with risk information, this reduced the variance in the WTP. The provision of information also reduced the WTP with the median WTP for the informed sample being approximately \$750-\$1,100 compared with \$800-\$1,900 in the uninformed group. This study demonstrated that the risk information did impact the WTP value with a tendency for individuals to revise their value downwards. It showed that uninformed patients are more likely to overestimate the risk associated with blood donations.

A third study attempted to evaluate the benefits of risk-free blood by measuring individuals WTP for haemoglobin solutions (Eastaugh 1991). Haemoglobin solutions have been produced to try and combat the risks associated with blood products such as hepatitis, human T-cell leukemia virus and human immuno-deficiency virus. The objective of the study was to establish if the consumers of hemoglobin solutions valued it more than (or equal to) the suggested retail price. The study sheds light on the limitations of planning a public service on the basis of preferences of individuals undergoing a catastrophe. This relates to the von Neumann-Morgenstern game theory that the higher the absolute risk, the more a person should be WTP per unit risk reduction (Neumann and Morgenstern, 1944). For example, most people would be WTP more to reduce the risk of death from 99% to 98% than for an equal percentage point reduction from 3% to 2%. The informed sample comprised 20 blood bank managers and 50 health services administration graduate students. The two groups were presented with five scenarios that varied the level of risk reduction from receiving hemoglobin solution. The sample were then asked how much they would be WTP to have hemoglobin solutions available, reducing their risk or eliminating their risk of infection from transfusion. The younger group (graduate students) had median WTPs that were lower than the managers over the age of forty. The authors speculate that blood bank managers are more educated in risk assessment and older therefore more in touch with their mortality. As expected, there comes a point in the sample, that the risk reduction is so high that the measured marginal benefit from each unit of risk reduction falls. In other words, as the risk reduction becomes greater the WTP for each unit of risk reduction diminishes.

3.10 Discussion

The market conditions of a publicly funded health care system like the UK NHS render it necessary to consider the economic consequences of making resource allocation decisions. This chapter discussed the various techniques that are available to conduct an economic evaluation. The difference between the approaches lies in the manner with which the 'effects' of the health technology are measured. With a CEA the benefits are measured in physical units, e.g. life years saved, number of successfully treated patients. A CUA incorporates the utility gained from an improvement from treatment and measures the outcome using QALYs as the unit of measurement. Finally, a CBA measures the effects of a treatment in monetary units using techniques such as 'the human capital approach' and the CV method.

A CEA is the most widely used form of analysis for the evaluation of health care technologies. However measuring outcomes in physical units means that it can be study-specific, e.g. blood cholesterol values, number of 'cancers detected'. This makes it difficult to compare cost-effectiveness ratios across different diseases. There also can be limitations concerning the categories of costs that have been included in a CEA study, and how the 'life years gained' measure may fail to take account of the morbidity of the patients. The CUA does take morbidity into account and measures the unit of effectiveness in QALYs. Also, having one unit of effectiveness means that outcomes can be compared across different diseases but the method used to calculate the QALYs can influence the final result.

There are other areas of health care that individuals may find important that are not measured by the tools employed in a CEA or a CUA. These aspects of health care have been discussed in this chapter. Individuals may find the process of care an important determinant of the value they place on treatment, also, an individuals attitude towards risk may have an effect on the perception of various treatments. Other aspects such as knowing a treatment exists should you need it (option value) and the value of information gained are important and the discussion in this chapter highlights how the CV method might encompass these benefits.

There are several reasons why the interest into the application of the CV method in health care has been growing over the last 10-20 years. There has been considerable progress in the development of the WTP elicitation methodology and in WTP estimation techniques. The fact that the CV method has the potential to capture wider aspects of health care such as the process of care, the option value and risk assessment make it a potentially attractive option.

Applying the CV method to health care however does have limitations. Considerable work is still required in establishing this approach as a reliable and valid means to assess health care benefits. Chapter four will provide a more thorough description of the methodological work that has been conducted on the CV method within health care. The discussion will describe the various issues of debate around the design issues and will discuss, among others, the choice of elicitation format, the appropriate instrumentation method, the biases inherent when using the method and overall, the methodological challenges of the technique will be outlined

Chapter Four – Methodological Considerations

4.1 Introduction

Over the years there has been a steady growth of interest into the use of the willingness to pay (WTP) technique within health technology assessment (O'Brien and Gafni 1996). In theory, there are three main arguments for using the technique as a research tool. First, as outlined in chapter two, it has a foundation in welfare economics as there are features of the technique that are 'theoretically correct'. Second, as explained in chapter three, the method provides the opportunity to consider aspects of health care that go beyond the traditional clinical gain. And third, since the benefits are measured in the same unit of currency as the costs, this allows decision-makers to pursue improvements in allocative efficiency (Donaldson, Farrar et al. 1997; Drummond, O'Brien et al. 1997).

Although the method has become more popular, it is far from refined and there remains many methodological questions that are unanswered in the literature. Examples of the sort of issues at debate are choice of the appropriate elicitation format, the amount and type of information to provide respondents and the choice of direction of measurement to be taken. This chapter is split into two parts; part A discusses the choice of direction of measurement and elicitation format; part B then focuses on the inherent biases and potential influences when using the method to evaluate health care interventions. The following CV design issues are each discussed in turn; validity, ordering effect, starting point bias/yea-saying, range bias, instrumentation technique, choice of elicitation format, embedding effect, the warm-glow effect and then finally, strategic bias. Throughout the chapter, examples of studies in the literature will be discussed and used as evidence towards resolving some of the methodological issues. Health care examples and where appropriate, environmental studies are both used to provide evidence. As well as these contributions towards the debate, the chapter will also look at the guidelines published on the use of the CV method in environmental economics, i.e. the National Oceanic Atmospheric Association (NOAA) guidelines.

4.2 Method for selecting papers

To identify the papers for discussion in this chapter and indeed throughout the thesis, the following criteria was applied. The focus was on papers published during the 1984-2003 period with an application of the CV method to health care. Particular relevant papers published in an environmental or experimental economics context were added to the discussion but the main focus was deliberately kept to a health care context. The papers were selected from a number of sources. Computer-based bibliographic databases were searched to identify papers written in English. Medline and EconLit were the two databases used. The following keywords were used: contingent valuation; willingness to pay; willingness to accept; cost-benefit analysis; cost-benefit evaluation and then these words were used in combination with: elicitation; bias; question type; question design; question format; effect; reliability; validity. In addition to the two databases, health economic conference proceedings were also searched using the above keywords as the search criteria, in particular, the proceedings from the UK 'Health Economics Study Group' and the 'International Health Economics Association' conferences.

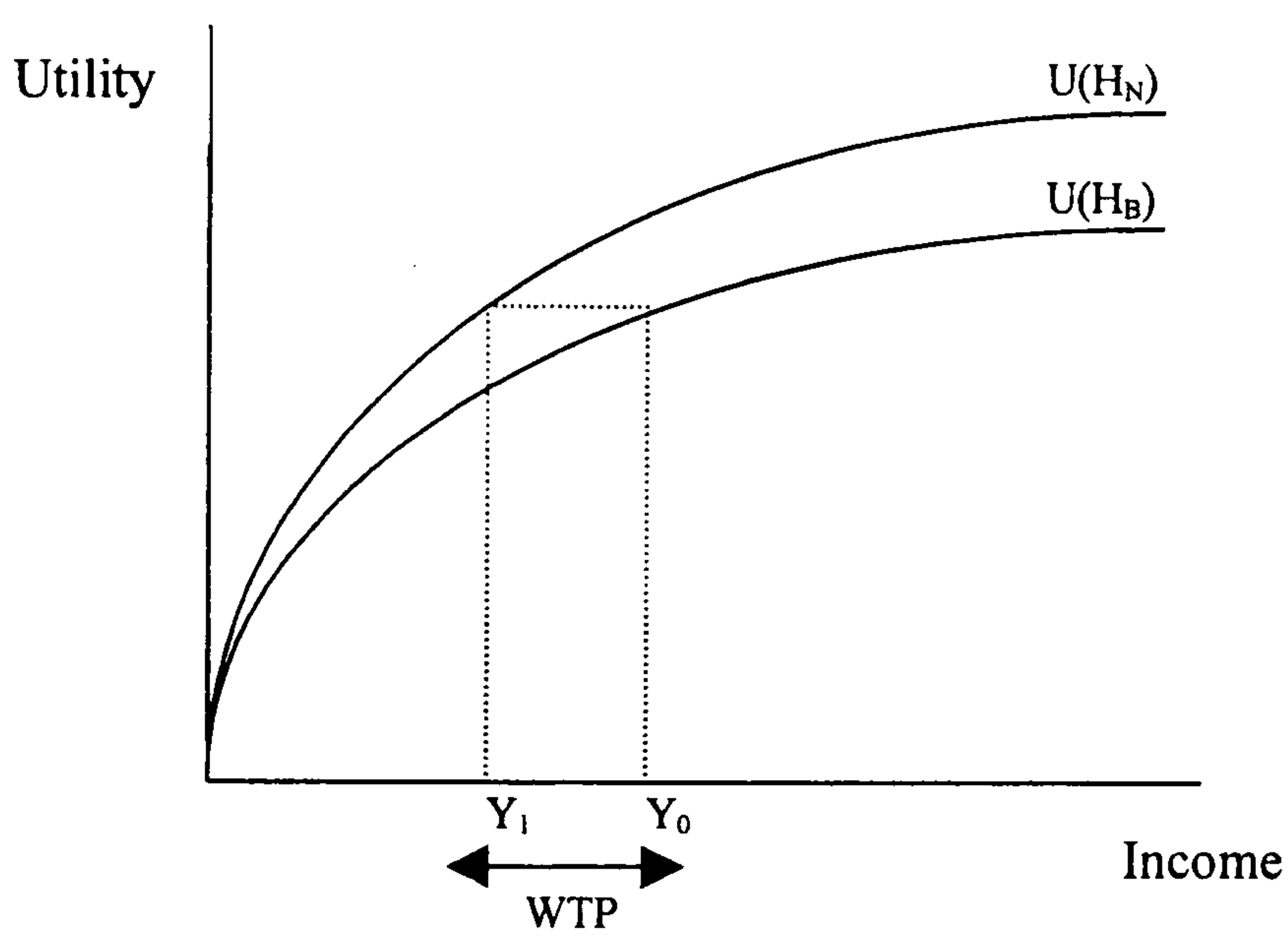
Through discussions on the methodological issues of the CV method, this chapter cites many studies that have been published since 1984 through to 2003. However this is not an exhaustive list of all studies and for this reason table A1.1 along with a short summary are provided in appendix one. This provides an overview of the methodological design elements and disease categories that have emerged as being popular areas of research.

Part A – Direction of Measurement and Elicitation Method.

4.3 Measurement direction: willingness to pay versus willingness to accept.

The effects of a health care programme can be measured in two different ways by either estimating the compensating variation or the equivalent variation. Both methods involve measuring the monetary amount required to keep utility levels constant, the compensating variation (CV) measures the amount starting from the original utility level while the equivalent variation (EV) measures it from the utility level after the change has taken place. The direction of payment, i.e. willingness to pay (WTP) or willingness to accept (WTA) depends on whether a welfare gain or loss has occurred. For example, if we imagine a proposed programme change that will result in a welfare gain then the compensating variation will be the WTP to ensure that the gain occurs. If the programme produced a welfare loss then the WTA value is the amount required in order to tolerate that loss. Put differently, the EV measures, for a welfare gain, the WTA to forego that gain or for a welfare loss, the WTP to prevent that loss. The distinction between the two forms of measurement lies in the manner with which the utility level is measured. The compensating variation is the money income adjustment necessary to keep the individual at his/her initial utility level throughout the change in provision while the EV is the money income adjustment necessary to maintain an individual at his/her final (post-change) utility level. This can be explained more clearly using a diagram (presented in figure 4.1) where we can assume that a health drug is introduced that improves the health state of an individual from having a high blood pressure (H_B) to having a normal blood pressure (normal health state defined as H_N).

Figure 4.1: Compensating Variation



If the utility level, before the change in health status is held constant, the WTP for the improvement in health from H_B to H_N is a measure of the compensating variation. The WTP in figure 4.1 therefore is the amount of money (Y_0-Y_1) that if paid will keep the individual at the initial utility level (the utility level with high blood pressure).

There are problems associated with estimating WTA values in practice as theory has shown that there is a tendency for individuals to overstate their WTA value. Evidence shows that WTA values for non-market goods are typically two to five times higher than the WTP values (Dubourg, Jones-Lee et al. 1994) and this disparity between

WTA and WTP is frequently reported in the literature (Knetsch and Sinden 1984; Brookshire and Coursey 1987; Coursey, Hovis et al. 1987). Potential reasons for the disparity are firstly, the endowment effect. This effect assumes that individuals are loss averse therefore once individuals own goods they value them more highly so therefore you would have to compensate them more than they would be WTP for them. Bateman et al state that 'all other things being equal, a change in consumption of a good X has a higher monetary equivalent when the change is seen as a loss, compared to when it is viewed as a gain' (Bateman, Longford et al. 2000). Secondly, the degree of substitutability is important, meaning that the easier it is to substitute the good that is valued, then the smaller the disparity (Hannemann 1991). Although the observed differences between WTA and WTP values have been widely researched outside the field of health economics, almost all research studies applied to a health care context choose to measure the compensating variation WTP value (Klose 1999).

However, among the few studies that have measured the equivalent variation, issues such as measuring heart patients' WTP for changes in angina symptoms (Chestnut, Keller et al. 1996) and the assessment of levels of utility between magnetic resonance (MR) and conventional angiography (Swan, Fryback et al. 1997) have been researched. To date, only one study has used the equivalent variation to measure a WTA value, in this case, it was for a community-based preventative programme (Lindholm, Rosen et al. 1994).

Dixon et al attempted to estimate the compensating variation WTA values. The study estimated the value of the loss in utility of potential losers from a fluoridation programme. The affects of a fluoridation programme were described to the study sample who were then asked if they would be in favour of the programme proceeding. If the individuals said 'no' they were either asked for their 'WTP to prevent' or asked if they would be WTA a tax rebate as compensation for the fluoridation programme to go ahead. The authors concluded that the 'WTP to prevent' questions were easier to answer than the 'WTA compensation' questions however the method performed no better in terms of the number of protests (to the style of question asked) generated. (Dixon and Shackley 1998).

Another study assessed the WTP and WTA from an insurance based perspective. By framing two scenarios, the study estimated the value of a new pharmaceutical (filgrastim). First, the drug was assumed not to be part of the insurance plan and the study sample were asked for their WTP to have the drug put on the plan. Second, the respondents were told that the drug was part of the plan and they were asked what their minimum reduction in premium would be to have the drug removed, i.e. their WTA. The results showed that the WTA amounts were significantly higher than the WTP values, in fact, the mean WTA ($47.6) was more than twice as great as the mean WTP ($20.3) for the same level of risk change (O'Brien, Goeree et al. 1998).

An experiment, based in the United States, set out to estimate the 'WTP to prevent' and the 'WTA as compensation' for a reduction in the Senior Companion Programme (SCP). The SCP organised elderly volunteers to visit elderly people to provide home/help and companionship. Sixty people were interviewed to find out how they felt about a reduction in this service either through their WTP to prevent a reduction or their WTA compensation for the reduction. The payment and compensation were in the context of social security payments and the sample were asked about two levels of reduction in the service (25% and 75%). In both cases the results showed that the WTA significantly exceeded the WTP (Garbacz and Thayer 1983).

It is rare for authors to explicitly describe their study as measuring either the compensating variation or the equivalent variation, however most studies do state whether they are measuring WTP or WTA. In a review of 42 papers by Diener et al. in 1998 it was found that only seven papers explicitly stated the utility change as being one of compensating variation or equivalent variation (Diener, O'Brien et al. 1998). Table 4.1 has been extracted from this paper and is a clear representation of the bias towards compensating variation WTP papers in health care.

Table 4.1 Direction of measurement in health care studies.

		No. studies		
		WTP	WTA	Total
Compensating Variation	Explicit	6	0	6
	Inferred	37	2	39
Equivalent Variation	Explicit	1	0	1
	Inferred	2	0	2

Source: A Diener et al. *Health Econ.* 7: 313-326 (1998)

Since most health care studies have adopted the CV WTP approach there is very little guidance on what is the most appropriate measurement to use in the health economics field. Most of the evidence does point towards a tendency for individuals to over estimate WTA values leading a lot of researchers to measure WTP values instead. Also, as will become clear later in this chapter it is vital that health care scenarios are presented to individuals in an understandable format and given that individuals are not accustomed to being compensated for unwanted health care changes a WTP scenario is much easier to communicate.

4.4 Elicitation Methods

When designing a contingent valuation study there are a number of elicitation formats to choose from. Each design has its own strengths and weaknesses and the debate is continuing regarding which method is superior. The next section of the chapter describes each elicitation format in turn, a more detailed section on the particular measurement aspects of each design will follow later in the chapter.

4.4.1 Open-ended question

The open-ended question is the 'simplest' of elicitation designs. The respondent is asked for their WTP for a health care good without any 'prompts' or interruptions from the questionnaire or interviewer. Usually the respondents are provided with a space (a line to write on) for their final maximum WTP value. An example of what an open-ended question looks like is as follows:

What is the maximum that you would be willing to pay to have (the health care good being evaluated) available to you?

Please write in the space provided. £______________.

In environmental economics, until the mid 1980s, most CV surveys used some version of an open-ended question (Hanemann 1994).

4.4.2 Iterative bidding technique

This elicitation format is termed the 'bidding game' and first originated in environmental economics having been introduced by Davis (Davis 1964). The question is designed so that it resembles an auction as the respondent enters a bargaining process with the interviewer. This process can be likened to a 'haggle' technique that happens in real-life markets making the process more familiar to the respondents. What happens is that the respondent is presented with a first bid and depending on whether they accept or reject it the bid is either raised or lowered till eventually the respondent's maximum WTP is reached. Some researchers claim that this technique is advantageous as the bidding process is more likely to measure consumer surplus (it will capture the highest price that consumers are WTP) (Cummings, Brookshire et al. 1986), and it is more likely that the respondent will give consideration to the value of the amenity (Hoehn and Randall 1983). Others reckon that it gives respondents more time to consider their preferences (Bateman, Willis et al.).

4.4.3 Payment scale question

The payment scale question design was developed by Mitchell and Carson in 1981 and 1984 as an alternative to the bidding game approach (Mitchell and Carson 1981; Mitchell and Carson 1984). The payment scale question presents respondents with a range of values to choose from. A typical design would ask the respondents what their maximum WTP would be for the good being evaluated and present them with a series of amounts, in a vertical list from the lowest bid (top) to the highest bids (bottom) in increments. If the respondent's maximum WTP went beyond the highest bid then they would be provided with a space to write down their maximum WTP value hence the question would default to an open-ended design. Respondents are asked to put a tick next to the amounts that they are sure they would pay, put a cross next to the amounts that they are sure they would not pay and circle their maximum WTP.

4.4.4 Closed ended/dichotomous choice/discrete question

Closed-ended questions are designed to lead to a yes/no response. The closed-ended technique was developed by Bishop and Heberlein (1979) (Bishop and Heberlein 1979). With this question design the respondents are presented with a bid and asked if they are WTP that amount and the bid levels are varied across the sample so that it is possible to calculate the percentage of respondents who are WTP as a function of the bid. Several motivations for using the dichotomous choice approach have been offered in the literature. The question design resembles a 'real-life' situation whereby the respondents are presented with a 'price' at which they decide whether or not they want to buy the good, as consumers are used to facing offers of goods at given prices, it seems a logical way to ask the question. Given it's simplistic nature, the closed-ended design is recommended by Mitchell and Carson as being suitable for a mail questionnaire study (Mitchell and Carson 1989). It is also claimed that the question design is less stressful to the respondent (Cameron 1988; Hanemann 1993). The first application of the closed ended approach was in a study concerning the use of goose-hunting permits (Bishop and Heberlein 1979) and it has become one of the most commonly used techniques in environmental economics (Johannesson and Jonsson 1991). As you do not get as much information from each respondent, the method does require a large sample size.

4.4.5 Closed ended with follow-up question

The closed ended with follow-up approach was proposed by Carson, Hanemann and Mitchell (Carson, Hanemann et al. 1986). This technique is an extension of the closed-ended (CE) method, to obtain more information from each respondent, a follow-up open-ended question is added. The design therefore is a form of bidding that is truncated at two bids. The use of additional follow up bids, or even triple-bounded discrete questions have also been explored (Langford, Bateman et al. 1996). Using a follow up question lessens the need for such a large sample size as you get more information from each respondent.

4.4.6 Contingent ranking question

This method requires alternative scenarios to be placed in order. The respondents are presented with cards each with a scenario representing various levels of improvement with corresponding bid amounts. Hence the respondent is paying different amounts for varying levels of improvement of symptoms. The respondent is then asked to rank the cards (alternatives) and the marginal WTP is calculated by the use of the discrete choice model which enables the ranked data to be analysed (Greene 1997).

Part B - Sources of Bias

Depending on the design of the contingent valuation study, there are inherent biases and influences that may occur. The following section of this chapter outlines each of these in detail. Throughout the discussion a broad range of issues are covered with reference made to published work as contributions to the debate.

The first methodological issue to be covered is the validity of the technique. Validity can be categorised into three themes, content, construct and criterion, each will be discussed in turn.

4.5 Content validity

The content validity of a study refers to the content of the survey instrument and related materials. If the valuation question creates incentives for the respondent to answer strategically, then this would reduce the content validity. Mitchell and Carson (1989, p.192) suggest that one should ask the following questions when evaluating the content validity of a CV study:

- *Does the description of the good and how it is to be paid for appear to be unambiguous?*
- *Is it likely to be meaningful to the respondents?*
- *Are the property rights and the market for the good defined in such a way that the respondents will accept the WTP format as plausible?*
- *Does the scenario appear to force reluctant respondents to come up with WTP amounts?*

One might argue that this is the most important type of validity to assess. If the content validity of a study is not correct then there would seem little point even debating the construct or criterion validity (Bishop, Champ et al. 1995). Content validity can be broken up into three different categories:

1. Context – the market or situation in which the good is being evaluated must be understood. The goal in designing a contingent valuation survey is to formulate it

around a specific commodity that captures what one seeks to value, yet is plausible and meaningful to the respondent (Hanemann 1994). When the good is a public one, as in the case of a tax-financed health good, it is recommended that the question should be asked in a referenda context as this is what respondents are familiar with as they might find the private good context artificial. To describe the valuation context to the respondent, three elements should be covered: (i) who the other participants are, (ii) whether the measure of value will be WTP or WTA, (iii) the value elicitation device (question format) (Bishop, Champ et al. 1995).

2. Payment mechanism - sometimes the payment vehicle used to elicit the WTP value can be misunderstood (payment vehicle bias). A full description of how the payment will be paid must be provided, i.e. direct payment, donation or through income taxation. The choice of the payment vehicle has to be one that respondents can easily understand and fit into the scenario being evaluated. The payment vehicle should not cause any undesirable reactions, i.e. respondents may object to being asked to pay more through income taxation to health care goods. Randall et al. found significant evidence of payment vehicle bias, but advises that careful instrument design and testing may be sufficient to alleviate most of the problem (Randall, Ives et al. 1974; Harris 1984). The timing of the payment structure is just as important as the vehicle as respondents need to know if the payment being made is a one-off payment, if so, when is it to be paid or, alternatively, will it be made in stages. To avoid misunderstandings the scenario should make this clear to the respondent.

3. The final category of content validity is the amenity validity – the good that is being valued or the quantity change in that good must be fully understood by the respondent. Misunderstandings may happen if the respondent perceives the good in a symbolic way and expresses a general attitude towards the good instead of a valuation for the good. To test for this, the researcher can vary the levels of change in the good to check that the WTP moves in the direction, as expected. Muller and Reutzel (1984) set out to test the impact on willingness to pay of changing the level of risk reduction on car crash protection. They stated the size of the risk reduction either on the base of 100 or on the base of 10,000. The resulting distributions and medians suggested greater willingness to pay when the level of risk reduction is expressed on the base of 10,000 rather than on the base of 100. On average, the respondents were WTP more for twice the level of risk reduction offered (Muller and Reutzel 1984). Another example of testing for this is when Johannesson investigated subjects' WTP for a reduction in incontinence symptoms. The study varied the level from 25% to 50% in the reduction of micturitions and leakages and also the number of leakages. They found that for each price asked (closed ended format) the proportion of subjects WTP that amount was higher for the 50% reduction compared to the 25% reduction. Therefore the WTP increased significantly with both the current level of micturitions and leakages, the size of the reduction in micturitions and leakages and with increased income. The author concludes that the magnitude of these significant variables seems reasonable and appears to support the validity of measuring the WTP of health changes using survey methods (Johannesson, O'Conor et al. 1997). Kartman tested for this scope effect when using a follow up question to a closed ended question. The study sampled four hundred respondents

diagnosed as having reflux oesophagitis who were asked for their WTP for successful treatment for the disease. To test for the scope effects, a split sample approach was used to evaluate different probabilities of successful short term treatment, and different probabilities of having a relapse with or without long term treatment. The author actually found that the closed ended question was sensitive to the scope effects but the variable testing for scope effects in the open ended follow up question failed to reach significance. The author concludes that these findings are in line with recent recommendations from an expert panel commissioned by NOAA that discourage the use of the open-ended question design format (Kartman, Stalhammar et al. 1997). These guidelines will be discussed in more detail in a later section of this chapter.

When communicating the scenario, visual aids are often used to help the respondent understand the information that is being provided. The individual must also be aware of the timing of the change being valued - this is because it is common to discount changes that are happening in the future so the individual needs to know if the change is happening now when the payment is made or at some time in the future (Bishop, Champ et al. 1995).

4.6 Construct validity

As well as the content validity the construct validity is important. To describe what is meant by construct validity the term can be divided into two categories; theoretical and convergent:

1. Theoretical – the scenario being described to the respondent should be correct within economic theory. One classic example of this is the relationship between WTP estimates and income as one would expect a positive correlation between the variable representing the level of income and the value of WTP expressed (assuming a normal good). If the coefficient on the income variable is negative this casts some doubt on the theoretical validity of the CV results (Mitchell and Carson 1989).

Lee et al. tested the impact of risk information on patients' willingness to pay for Autologous Blood Donations (ABD) (Lee, Liljas et al. 1998). *A priori*, they expected that the greater the perceived likelihood and risks of allogenic transfusion, the higher a patient's WTP for ABD. The results were consistent with the theoretical predictions about the impact of risk information as the WTP was highly related to the perceived risk of requiring a transfusion, income and perceived dread of allogenic transfusions.

Kartman et al. assessed the theoretical validity of the CV method by assessing if the WTP varied for reductions in angina pectoris attacks (Kartman, Andersson et al. 1996). The study varied the percentage reductions in anginal attacks randomly in different subsamples. Data were collected about angina pectoris status, attack rate and income to test the validity of the CV method. It was found that the WTP was related to income, angina pectoris status, attack rate and percentage reduction in attack rate in the direction that one would expect. The mean WTP amounted to 1,145 SEK for the 25% reduction and 2,780 SEK for the 50% reduction in attack rate.

2. Convergent – this type of validity is achieved when the measure in question and a second measure 'converge' in a manner predicted by theory. This can be measured by assessing the values elicited by a contingent valuation method compared to values estimated through alternative outcome measures such as the time trade off technique or the standard gamble approach.

4.7 Criterion validity

Mitchell and Carson (1989) say that in order to evaluate the criterion validity of a contingent valuation estimate it is 'necessary to have in hand a criterion which is unequivocally closer to the theoretical construct than the measure whose validity is being assessed' (Mitchell and Carson 1989). Bishop (1995) suggests that actual market prices would be one way of doing this but unfortunately actual market prices do not exist in a health care market that is tax-financed. Alternatively, respondents can be asked for a hypothetical WTP value and then provided with an actual opportunity of purchasing that good at a later date, but again, this is not really practical in a tax-financed health care system. Bishop argues that the best way to test for criterion validity is through simulated markets. These markets would involve creating situations (in the field or laboratory) where the respondent has the opportunity to actually pay for the good. Such markets should be fully parallel to actual transactions taking place. Therefore simulated markets should provide values that are closely related to true CV values. Kealy et al state that:

The WTP values measured in a simulated market are the best available criterion to evaluate the self-reports of WTP from the corresponding hypothetical situation posed by the contingent valuation method (Kealy, Montgomery et al. 1990).

Hanemann also argues that replication is a useful method of validating responses, at least on the small scale (Hanemann 1994). It is a good way of checking if the results are stable and the instrument is being communicated as intended.

In environmental economics, it is possible to make comparisons with estimates obtained through indirect methods such as the travel cost method, and there are over 80 studies that do this. WTP estimates were compared to travel cost estimates in a study that investigated improving access to mammographic screening in rural areas in Australia. This study found that the WTP estimates ($148.09) for the use of mobile screening units were significantly higher than the travel cost estimates ($83.10). The author suggests that this might be due to the WTP estimate capturing the altruistic benefits of mobile screening but suggests that the difference may be due to a range of potential biases with both methods (Clarke 2002). Overall, the studies find that the contingent valuation estimates are fairly close to the indirect estimates. The contingent valuation estimates are usually slightly lower than the revealed preference estimates but they are normally highly correlated with them (Carson and al 1994).

4.8 Question Order

Theoretically speaking the response to a CV question should not be influenced by the question order in the survey. The same response should be given if the CV question is asked by itself or with other questions. Kartman et al conducted a study that tested the question order effect by assigning respondents to different question order sequences. Willingness to pay data for superior treatment of reflux oesophagitis was obtained. Out of a sample of 400 respondents, no significant question order effects were detected in the willingness to pay data. The authors do mention however that this result should be treated with caution as the same bid was suggested in all three questions, and that the different questions that were varied concerned different health commodities (Kartman, Stalhammar et al. 1996). The effect of the question order may be more important in a study that uses different bids in different questions as this may cause the respondent to anchor their bid from the first question for all subsequent questions. Stewart et al examined the possibility of ordering effects and found them to be observed in the ranking of programmes, in the proportion of zero values reported and in the WTP for one of the programmes. These authors explain the ordering effect as one of 'fading glow', whereby the first programme in any sequence captures much of the utility associated with giving (Stewart, O'Shea et al. 2002).

4.9 Value cues

Another issue around the design on contingent valuation studies concerns the potential for 'value cues'. The design of the question may provide respondents with a 'cue' as to what the expected response is to be. As with the validity issues, 'value cues' can be divided into two sections: psychological anchoring effects and range bias.

4.9.1 Psychological anchoring effects

The anchoring effect is a potential source of bias within contingent valuation studies. Some researchers even feel that the effects present more of a significant problem for contingent valuation studies than strategic bias (O'Conor, Johannesson et al. 1999).

One such psychological anchoring effect is defined as starting point bias and is inherent in studies that adopt the iterative bidding technique as the elicitation format. This is because the final WTP value may be influenced by the starting bid used in the question. To detect this, the starting bid can be varied across respondents to establish if those who start at high bids give significantly higher WTP values compared to a sub-group of respondents who start on low bids. The evidence for starting point bias is equivocal; some studies have found it exists, other have not (Stalhammer 1996). Liu et al. used the contingent valuation method to estimate mothers' WTP to protect themselves and their children from suffering a minor illness in Taiwan. WTP was elicited using a binary choice question design where each respondent was assigned to one of three initial bids (NT$300, 700, 1000). Results suggested that the effect of starting point bias was modest, the estimated WTP was about 25% smaller for both mother's and child's cold using the smallest rather than the largest bid (Liu, Hammitt et al. 2000). O'Brien et al used a 2 x 2 factorial study design to estimate the effect of starting point bias in a study that assessed the use of the contingent valuation method to measure the monetary value of a new drug as prophylaxis against febrile neutropenia after chemotherapy treatment for cancer. The 2x2 factorial design assigned respondents to two different bidding algorithms and dollar starting points (the study also varied the size of the treatment effect). The results suggested that there was no significant starting point bias when the different bidding algorithms were compared (O'Brien, Goeree et al. 1998). Another study that O'Brien conducted with Viramontes directly tested for starting point bias by directly comparing the WTP method with alternative means of eliciting preferences. WTP was elicited by a single bidding game where the respondents were randomly assigned to one of five starting bids. After adjusting for income and health status the authors found no association between starting bid and mean WTP and thus concluded that there was no evidence of starting point bias (O'Brien and Viramontes 1994).

Starting point bias has been found to exist in other studies; Cho-Min-Naing et al. used WTP to estimate the demand for the ICT Malaria test kit in Myanmar and found evidence of starting point bias when respondents were allocated to a different starting bid. The higher the amount of the initial bid, the higher the mean value of WTP (Cho-Min-Naing, Lertmaharit et al. 2000). Dalmau-Matarrodona estimated the determinants of the WTP for home care services in day case surgery and found a clear presence of starting point bias (Dalmau-Matarrodona 2001). Eastaugh investigated the WTP in treatment of bleeding disorders and randomised half of his sample to receive a low bid questionnaire and the other half to receive a high bid questionnaire. Results showed evidence of starting point bias as if the initial bid started low, at $500, the median WTP response was $1,500. If the initial bid started high, at $5,000, the median WTP was $3,500 (Eastaugh 2000). Some authors have suggested design techniques to overcome the starting point bias problem. Reardon et al. suggests that randomly-generated initial bids administered across the study sample may be helpful (Reardon and Pathak 1989), Eastaugh argues that one can 'debias' the results by randomly assigning 10% of the study sample to one-tenth of the starting bids (Eastaugh 2000).

Within a closed ended format, an extreme form of starting point bias is referred to as 'yea-saying' by Mitchell and Carson (1989). This happens when respondents agree to pay regardless of the bid level offered. Many studies have found the final mean WTP value to be much higher with dichotomous choice questions than with open-ended question designs suggesting that 'yea-saying' is a problem (Boyle, Bishop et al. 1985; Seller, Stoll et al. 1985; Kealy and Turner 1993; Brown, Champ et al. 1996).

4.9.2 Range bias

Another type of value cue is range bias, usually encountered in payment scale question designs. Range bias is similar to starting point bias only instead of being influenced by the starting point bid the respondent is influenced by the range of values chosen for the payment scale design. Usually respondents are presented with a range of values starting from the lowest (usually £0) to the highest (usually £1000), if the study designs adopts a different range, i.e. £10 - £100, then a change in the overall WTP result is likely. It is also plausible that respondents may be sensitive to the positioning of the values within the range. The extent of range bias has rarely been investigated in health care (Klose 1999). The literature search found one paper that investigated this effect directly; Johannesson (1991) asked a group of respondents for their WTP for treatment of hypertension using two different payment card ranges and it was found that the broader range induced a higher, but not significantly higher WTP (Johannesson, Jonsson et al. 1991).

4.10 Instrumentation technique

As well as the validity issues and the potential threat of 'value cue' bias, the choice of the appropriate instrumentation technique is discussed a great deal in the contingent valuation literature (Reardon and Pathak 1989). To elicit CV amounts, direct interviews are particularly beneficial if the good or service being valued is difficult to communicate. When using a bidding game approach as the elicitation technique a personal or telephone interview is the recommended means by which to administer the survey. Obviously telephone interviews are the most economical way of designing the interview study both in time and in money. Smith conducted a telephone survey in 1983 that aimed to assess consumers' willingness to pay for pharmacists' clinical services. No instrument problems were noted in the study although the methodological discussion was limited meaning that it is difficult to draw conclusions about the instrumentation method for CV research (Smith 1983). Mail surveys are by far the most straightforward method of administering a CV study. When adopting a mail survey, the open-ended, payment scale or closed ended elicitation technique can be adopted. Limitations of a mail survey are the likely low response rates to the WTP questions and the limited scope to describe anything in detail. If a respondent fails to understand the question then the incentive is to leave it blank meaning that the rate of 'no response' to the WTP question is inclined to be much higher when using mail questionnaires. Response rates for mail surveys range considerably from being as high as 99% (Keith, Haddon et al. 2000) to 51% (Donaldson, Shackley et al. 1997). According to Bishop et al. the choice of the 'best' instrumentation technique involves trading off the limitations and advantages of each option:

Some would argue that personal interviews yield more accurate results because of greater flexibility [than mail survey] in the amount and nature of the information that can be provided to subjects and because of opportunities for iterative bidding. Interviews may in fact provide opportunities to help subjects more fully explore preferences and constraints to predict more accurately how they would behave in real markets. Our concern is that such interview procedures would be superimposed on inherently artificial contingent markets. Doing so may cause *subjects to base answers even more on the information received and other aspects of the interview situation and less on the relevant economic parameters. Only further empirical work can determine which basic approach is correct (Bishop, Heberlein et al. 1983).*

4.11 Appropriate question format

As already discussed, the choice of elicitation method is dependant upon the choice of instrumentation technique. Certain question formats are better suited to instrumentation techniques than others. For example, the iterative bidding question design involves a given amount of interaction between the respondent and the researcher making a personal or a telephone interview the most appropriate setting. Each elicitation design comes with its own strengths and limitations and the debate concerning the most appropriate format is far from resolved.

Within a health care context, there have been few studies that have focused on comparing the performance of different elicitation methods. Donaldson et al. (1997) proposed that the open-ended question design led more respondents attempting to estimate the 'resource cost' of the good or the programme being evaluated. As mentioned earlier, the open-ended question provides respondents with no prompts or cues and they are given no guidance on how to reach the maximum WTP value (Donaldson, Shackley et al. 1997). With this elicitation format Donaldson provides three explanations as to why respondents may consider the 'cost' rather than the 'value' of the good.

1. Respondents may become confused between the concepts of 'cost' and 'value'. When asked to value two alternative treatments, despite preferring the lower cost option, the respondent may become confused about the task and place a higher WTP value on their least preferred option (as they believe that to have a greater resource cost). Respondents can find it difficult to value the differences between alternative treatments when they focus on the perceived resource cost.

2. Due to the lack of guidance with open-ended questions respondents may reveal an 'acceptable' WTP amount (based on perceived resource cost) rather than the genuine 'maximum' value. The design of the question does not adequately explain what the study is trying to achieve.

3. The respondent may take on a 'social duty' and explicitly choose the alternative treatment option that they believe will be the least costly to the NHS.

The role of 'cost' was examined in a study by Baron and Maxwell that was designed to investigate the influence of information on cost (and benefits) (Baron and Maxell 1996). The authors asked WTP questions concerning hypothetical public goods and found that respondents were willing to pay more for goods that are more expensive, holding benefits constant. They used an open-ended question design with a referendum format and randomly allocated different scenarios by varying the costs and benefits across the sample of respondents. The authors found that cost affected WTP quite substantially. When the authors took out any cost information provided in the scenario, a situation similar to most contingent valuation studies, they found that the WTP was still affected but this time by 'implied cost'. The respondents were provided with quantitative information from which the size of the cost could be inferred. The authors argue that cost is a natural means of assessing WTP for a good as the respondents use cost to provide information about value, as a guide to what others are WTP and because it may be unfair to pay less than the cost plus a fair profit. Even when respondents are given information on the benefits hence they no longer need information on cost to judge benefits, they still consider the costs. The authors suggest taking out all information (i.e. means of provision of good) that can lead to costs being inferred and focus more on the benefits to be provided rather than the means of providing them.

The role of 'cost' is part of everyday life. People consider the costs of goods when they are making purchases in an actual market. The maximum willingness to pay for a good can incorporate additional factors over and above the perceived value of the good. For example, studies of supermarket sales show that shoppers are less likely to buy an item at a particular price if they believe that the usual price was lower (Winer 1986; Mayhew and Winer 1992; Rajendran and Tellis 1994). The shoppers consider past prices of the good and other similar goods and use this information to try to assess how the current item compares. In consumer behavioural research this is what is termed 'reference prices'. There is no reason to expect subjects not to go through the same thought processes when considering the maximum WTP for health care programmes. Given that in the UK the health market is tax financed, the respondents are limited with information on actual prices of health goods but it is likely that they will attempt to 'guess' the cost based on the information provided in the scenario description.

Another study that looked at the performance of different elicitation designs compared the open-ended with the payment scale technique (Donaldson, Thomas et al. 1995). The study compared the response rate, completion rate, the association between WTP and ability to pay (ATP), the mean and median values and the R^2 in regression analysis of WTP to establish which question format 'performed' better. A questionnaire design was adopted and overall, 380 questionnaires were returned giving a response rate of 61%. The overall response rate was made up of a response rate of 65% for the open-ended design compared to 58% for the payment scale design, this difference was not statistically significant at the 5% level. However a statistically different response rate (at the 1% level) in the WTP question, amongst the questionnaires returned, was detected with more people answering the payment scale question. The mean and median WTP in the payment scale group was higher and consequently the authors suggest that the payment scale question is therefore more likely to reveal true consumer surplus. The distribution of WTP across social class groupings in the sample were compared for both question designs to detect the association between ATP and WTP. The results showed that the payment scale group behaved as one would expect *a priori* with the WTP values getting smaller with the direction of social class from I through to V. The open-ended results did not display a constant falling WTP value across the social class grouping. The authors conclude that the results strengthen the case for the payment scale design as being a more valid approach to elicit WTP valuations but suggest that more work is needed into comparing the payment scale approach with the alternative of a closed-ended design.

Mitchell and Carson claim that the open-ended format leads to an unacceptably large number of non-responses or protest zero responses to the WTP question (Mitchell and

Carson 1989). Johannesson et al. in 1991, in a study that investigated the WTP for antihypertensive therapy also criticised the open-ended approach as resulting in low response rates (Johannesson, Jonsson et al. 1991). However, Narbro and Sjöström used the open-ended technique to examine the WTP for obesity treatment and found that the question design led to a response rate of 88% which they claim is equal to (Kartman, Stalhammar et al. 1996; Johannesson, O'Conor et al. 1997), or higher (Neumann and Johannesson 1994; Donaldson and Shackley 1997), than response rates of other question formats (Narbro and Sjostrom 2000).

Slothuus and Brooks compared the contingent ranking method with the double-bounded (closed-ended) method (Slothuus and Brooks 2000) in a study that asked one hundred and fifteen patients with rheumatoid arthritis for their WTP for the alleviation of symptoms using a novel antirheumatic agent (cA2). The authors found that as the bid levels increased the proportion of subjects responding with 'yes' decreased. There were no statistical differences in results between the two approaches of asking WTP.

In 1998, Olsen et al. asked a community for their views on three different public sector health care programmes, by eliciting their WTP values (Olsen and Donaldson 1998). The results from this study illustrate that not only is the choice of elicitation format important but also the style in which the question is framed. The three programmes valued were a helicopter ambulance service, more heart operations and more hip replacements. The study design used a referendum approach and asked respondents for their willingness to contribute in terms of extra earmarked taxation per annum. Respondents were asked to elicit WTP through a payment card and then asked an open-ended question for their reasons why they chose that WTP value.

Following this, the interviewer then read out some pre-coded reasons and asked the respondent to indicate on a visual scale which reasons were important. Respondents who were averse to paying by taxation were asked if they would like to make a voluntary donation. The results from 143 interviews were analysed using Ordinary Least Squares (OLS) regression analysis to test for factors associated with WTP for each of the three programmes. The 'open ended' reasons provided for WTP were grouped and entered as dummies into the regression and were found to have no influence on WTP. The authors state that the pre-coded reasons proved more 'fruitful' in explaining WTP. The respondents were asked for their valuations for the helicopter ambulance service, more heart operations and then hip replacements in that order. Lower WTP values were given for the hip replacement operations suggesting that this may be attributed to the presence of an ordering effect. An attempt was made to reduce the possibility of this happening by reading back the ranking of the programmes and giving the respondents the opportunity to change their WTP values. Twenty-two in the sample of 143 took this opportunity and changed their values. This study raises an important issue concerning the framing of the elicitation design. Given that most health care services are publicly financed the study frames the WTP question in the context of contributing to community health services. By asking individuals for their 'willingness to contribute' instead of 'willingness to pay' the question becomes one of valuing a community program where the benefit is to everyone and not just to the individual person doing the valuing. However this type of framing may increase propensity to overstate values due to the effect of altruism and the authors do recognise that 'community bias' (have to be seen to be doing the right thing) may have influenced the results.

Ramsey et al. demonstrated the difficulty in using a bidding game approach when the study design used a mail survey. The study investigated the WTP for antihypertensive care. Mail questionnaires were issued with a WTP question that contained 10 separate WTP 'bids' ranging from $25 to $250, in $25 increments. They found that even at the highest bid, 24% of the sample responded with a 'yes' result and given that the design was a mail questionnaire, there was no scope to investigate this further (Ramsey, Sullivan et al. 1997).

When a comparison was made between the closed-ended and the closed-ended with a follow-up question (for the WTP for alleviation of arthritis symptoms). The authors found that the two question designs produced significantly different WTP amounts (DKK 637 v's DKK 1,268). They suggest that the inclusion of a follow-up question increases the precision of the results. The authors recommend that when choosing between the two question designs, power, efficiency and size should be used as a selection criteria (Slothuus, Larsen et al. 2000).

4.12 Whose values count?

Given that the contingent valuation method is part of a cost-benefit analysis that takes on a societal perspective, the WTP question should be presented to all subjects that stand to lose or benefit from the proposed programme. If this sample is too large, then a representative sample will be appropriate. By asking not only subjects that benefit directly from the treatment but also subjects that benefit indirectly, the study should then in theory, incorporate any 'externalities' produced.

It may be easier however to assess the actual patients in a CV study (Neumann and Johannesson 1994; O'Brien and Viramontes 1994). The benefits of a programme can then be measured by multiplying this WTP with the probability of becoming a patient. Patients are expected to be more fully informed about the potential health consequences of treatments and since they are already in the health state there is no need to inform them of the probability of needing a specific health service (Johannesson 1996). Most studies use patients as a sample as they are a convenient population to capture (Donaldson and Shackley 1997; Donaldson, Shackley et al. 1997; Johannesson, O'Conor et al. 1997; Keith, Haddon et al. 2000; Dalmau-Matarrodona 2001). However Ortega et al. decided to use both patients and the general population in a study that assessed the WTP for erythropoietin in the prevention of chemotherapy-induced anemia. By including both patients and members of the general population, the study addressed the cost-benefit of erythropoietin from both perspectives. Patients were asked how much they would pay for the drug and to make it realistic to the general population, the payment for the drug was represented as an increase in health insurance premiums .

The question of whose values count can be resolved by examining the context of the health care decision. In a situation where the health care budget is fixed and patients prefer an alternative care that is more expensive than the status quo then more resources would have to be obtained from elsewhere to implement the preferred care. In this case, the decision makers can either use the values of different patient groups (if available) or ask the community on their values on how the health care budget should best be allocated.

4.13 Embedding effect

The embedding effect was first analysed systematically by Kahneman and Knetsch in 1992 (Kahneman and Knetsch 1992). The effect happens when the respondent finds it difficult to isolate a specific case from overall considerations when they are deciding on their maximum WTP value. For example, if the respondent was asked to consider two alternative treatments for the same disease and they gave the same WTP value for both, they may be considering the treatment of the disease generally and not considering the different aspects of the two alternative treatments. The embedding effect has been detected in environmental economics when respondents allocate their full environmental account to one specific environmental issue, i.e. they fail to consider the opportunity cost of the amount they are claiming to be WTP (Kneese 1984). The willingness to pay responses are highly similar across different surveys, even where theory suggests that the responses be very different. Mitchell and Carson refer to this when 'respondents react to an amenity's general symbolic meaning instead of to the specific levels of provision described....a propensity to respond to the symbol rather than to the substance' (Mitchell and Carson 1989). One possible way to overcome this problem is to ask respondents a follow up question, asking them if they would be willing to support other important issues with a similar WTP amount. Schkade and Payne did this when asking individuals for their WTP to protect migratory waterfowl. The authors confronted the respondents by asking them if they would be WTP for other good causes. The respondents suddenly realised the far-reaching implications for their household budget of their previous WTP response, and indicated that the amount they stated was really too large or that it should go for all similar issues (Schkade and Payne 1994).

The same affect could happen in health care when respondents are asked a follow-up question to the initial WTP question. This follow-up question could ask for the WTP for an alternative health care treatment that could substitute the treatment valued in the first question. Respondents then will consider the total value of the WTP amount stated in the first question in relation to their total WTP for treatment of the disease.

4.14 'Warm glow' feeling

People may respond to the contingent valuation question in a manner that does not reflect their true preferences. One such example is that the individuals may receive a 'warm glow' from expressing support for a good cause (Andreoni 1989), respondents may get some kind of moral satisfaction when deciding upon their WTP response (Andreoni 1990). This theory is based on the 'impure altruism' model that individuals contribute towards a good for two reasons; first, they simply demand more of the good and second, individuals gain benefit from their contribution per se, like a warm-glow. Andreoni has proved the model of impure altruism to be a straightforward but powerful predictor of the act of giving and that it is consistent with empirical observations (Andreoni 1989). Individuals demand more of the good for altruistic reasons, they will also gain benefit from the act of giving (warm glow), this results in an 'impure altruistic' motive. It is reasonable to assume that preferences will include a combination of both the altruistic and the warm-glow effect as individuals actually care about the good but also contribute to receive a warm glow as well. Alternatively, the respondent may feel some moral obligation towards the scenario being valued and may give a WTP value so that they feel they are 'doing their fair share' (Diamond and Hausman 1992). When Schkade and Payne (1993) conducted the analysis on subjects' willingness to pay to protect migratory waterfowl, the authors asked the subjects to 'think aloud' to try and understand the thought process that the subject went through before deciding on a WTP value (Schkade and Payne 1993). The study found that one sixth of the sample made comparisons with donations to charities as a way of deciding on their maximum willingness to pay. Instead of the willingness to pay amounts reflecting personal preferences, the amounts were more likely to be consistent with the warm glow hypothesis.

The 'warm glow' feeling is closely related to the embedding effect. If respondents get pleasure from contributing to health care programmes to help the community, the willingness to pay for this warm glow is not determined by the *particular* health care project in question. The same WTP amount will be given for different types of health care programmes, regardless of what it is, if it helps the community. The respondents will give a WTP value for the warm glow feeling, not the health care programme. This theory is supported by studies that have looked at respondents WTP for several programmes simultaneously compared to their WTP for the programmes separately. No such study has been undertaken in the health care field. In environmental economics, Kemp and Maxwell (1993) asked respondents to state a WTP to minimise the risk of oil spills off the coast of Alaska, they calculated a mean WTP of $85. A different sample of respondents were asked for their WTP for several environmental programmes where they were asked to divide and subdivide their WTP among the programmes. By the time they got to the Alaska oil spill project the mean WTP had dropped to $0.29 (Kemp and Maxwell 1993). It seems that the respondents were expressing a warm glow effect with the desire to support environmental projects regardless of what they are.

4.15 Strategic bias

Strategic bias involves the respondent deliberately overstating or understating their true value. If the respondent feels that the good will be implemented if they give a high value but this will not affect how much they have to pay for the good, then they have an incentive to overstate the true value. However if they believe that the good will be implemented but the amount they will actually have to pay depends on their stated value, there is an incentive to understate their WTP value. Strategic bias is related to the 'free-rider' problem in economics. Within the UK NHS system, since respondents do not have to pay for health care directly, there is more scope for their WTP values to be revised upwards. Respondents realise that the health care good will be implemented if they give a high value but this will not affect how much they have to pay. Given the hypothetical nature of the WTP response there are no incentives for true preference revelation. However authors such as Hoehn and Randall (1987) argue that the dichotomous choice format provides an incentive for individuals to eschew strategic behaviour and respond truthfully (Hoehn and Randall 1987). There is little empirical evidence to support the hypothesis of strategic bias in CV studies (Bohm 1972; Smith 1979; Milon 1989).

Protest responses can be classified as strategic bias as the respondent 'protests' to the process of investigation by stating either a zero response or an unreasonably high or low response. Respondents may state no WTP, even though they care about the programme, as they feel that it is someone else's responsibility to pay, i.e. the NHS. It is difficult to detect protesters in a CV study, that uses mail questionnaires, as zero responses can be classed as genuine zero WTP values. One way of getting closer to the 'thought process' behind the WTP value is to ask respondents for their reasons why they chose that value. This may help to separate the respondents who are protesting to the method from the respondents who are giving genuine zero responses. Diamond and Hausman (1994) state that it is standard practice to eliminate 'protest zeros' as this type of zero is not a credible answer. They argue that if the respondent has answered all other questions in a manner that indicates that they do put a positive value on the changes in the level of the public good, but they have responded with a zero value to the contingent valuation question, then they should be removed from further analysis (Diamond and Hausman 1994).

In the Donaldson and Shackley study that used the WTP technique to measure the value of laparoscopic versus conventional cholecystectomy, one hundred and seventeen patients returned the questionnaire giving a response rate of 77%. Fourteen (12%) were not willing to pay anything for laparoscopic repair rather than conventional repair. Four of these 14 were people who explicitly registered a 'protest' and the authors dropped them from subsequent analysis (Donaldson and Shackley 1997).

An alternative method, if adopting an interview study, is to ask respondents to 'think aloud' and report everything that goes through their minds as they respond to the questionnaire. The thoughts are recorded on audio tape that can then be transcribed and coded. The researcher can then analyse the thought process of the respondents when trying to decide on a WTP amount and identify respondents who are not being consistent with theory who can then be removed from further analysis.

Dalmau-Matarrodona attempted to distinguish the zeros generated by economic decisions (true zeros) to those made by non-economic decisions (protest zeros). The author in this study argues that it may not be correct, statistically, to remove all respondents who have stated a protest zero. There is a potential cost to the loss of information and these respondents may not have the same characteristics as the rest of the sample therefore you may inflict sample bias. This study adopted a different type of model (Double hurdle model) to allow the identification of the mechanisms behind a zero response (Dalmau-Matarrodona 2001).

4.16 Preference Formation

A central assumption in the theory of consumer behaviour is that, given a feasible set of consumption bundles, the rational consumer will choose the bundle that he prefers most, or in economic terms yields the most utility – optimisation. This assumption that all rational people can choose amongst a set of options and therefore maximise their utility has some controversial associations (Heap, Hollis et al. 1992). It is plausible under conditions of perfect information as each individual will have knowledge as to how much utility will be gained under each bundle of goods. In the real world however individuals make choices under conditions of imperfect information. This uncertainty results in the need to generalise and instead of the individuals maximising their utility, they maximise expected utility. When individuals have to make decisions under conditions of uncertainty they assign probabilities to the outcome of each event. Based on the individual's knowledge and judgement, a probability value will be attached to the outcome of each event resulting in a probability distribution for all events. Expected utility theory (Von Neumann-

Morgenstern 1944) is based on predictions and reasonings about rationality; rather than on observing behaviour (Heap, Hollis et al. 1992). However people do not always act in a 'rational manner', a classic case of acting irrationally are when individuals show signs of preference reversal. The psychologists Lichtenstein and Slovic (1971) and Lindman (1971) were the first to observe preference reversal (Lichtenstein and Slovic 1971; Lindman 1971). Psychologists and economists explain preference reversal using two separate theories. The economists believe that individual behaviour is governed by *context-free* preferences where as the psychologists assume that the preferences are *context-sensitive* (Cubitt, Munro et al. 2000). A classic example of preference reversal concerns decisions relating to pairs of simple monetary gambles such as P- and $ - bets. The P-bet offers a relatively large chance of a modest prize while the $-bet offers a smaller chance of a larger prize. When the individual evaluates each of these alternatives the classic tendency is for the individual to indicate a preference for the P-bet but then place a higher monetary value on the $-bet. This is an example of a standard preference reversal and is illustrated in example 4.1.

Example 4.1

Consider these two gambles:

Gamble Q: 7/36 chance of winning 9.00, 29/36 of losing 0.50

Gamble R: 29/36 chance of winning 2.00, 7/36 chance of losing 1.00

Q and R are gambles on the number of a ball, to be drawn at random from a bingo cage. The balls are numbered 1-36, with prizes paid on the higher-numbered balls. Thus, for example, gamble Q produces a win of 9.00 win on balls 8-36 and a loss of 0.50 on balls 1-7.

Your first problem is to say which of the two gambles you would prefer to play, if given a choice between the two. Your next problem (which, in an experiment, would be given to you after some interval of time) is to say how much gamble Q is worth to you. Imagine that you have the right to play gamble Q, and state the smallest sum of money which you would be willing to accept in exchange for the gamble. Your final problem is to say how much gamble R is worth to you.

In experiments of this kind, many people choose R in the straight choice, but put a higher money value on Q. The straight choice of such people appear to reveal a preference for R over Q, while their valuations appear to reveal the opposite preference – hence the term 'preference reversal'. (This example given is taken from an experiment carried out by Grether and Plott (1979), with pay-offs in US dollars. There were 46 people in the experiment, each of whom was presented with six pairs of bets on the model of Q and R. This gave 276 observations. In 91 cases out of 276, the preferences revealed in straight choices were in the opposite direction to those revealed in valuations. Of these 91 preference reversals, 69 took the form in which R was chosen in the straight choice, but Q was given the higher valuation.)

Source: Hargreaves Heap et al. (1992) pp.43

4.16.1 Economic explanations for preference reversal

Loomes and Sudgen (1982) first proposed regret theory as an explanation of preference reversal. This theory predicts that preferences can be systematically non-transitive (P$\succ$\$ in choice and then \$$\succ$P in valuation, which gives the non-transitive ordering that P≻$~$≻P). Lets take an example where there is a probability (<1) of state Sj occurring. The individual will choose between A and B so as to minimise regret (or, equivalently, maximise rejoicing). If state of the world Sj occurred and Xa>Xb, this would entail choosing option A because 'what is' (Xa) provides higher utility than 'what might have been' (Xb). Loomes et al. developed an empirical experiment to test the predictive power of regret theory of choice (Loomes, Starmer and Sugden, 1989). Three pairwise choices were constructed, involving lotteries analogous to a P-bet, a $ bet, the third option guaranteeing of a given sum of money (C-choice). The experiment was set up as a choice-only design therefore the results from the experiment only provide explanations of choice and cannot explain differences between choice and valuation tasks. As explained in figure 4.1, the 'standard' preference reversal theory predicts that for an individual to reveal a preference reversal the individual's valuation of the two bets (P and $) are not in line with their preferences for the two bets. In the experiment that Loomes et al conducted, the results show that individuals choose the P bet over the $ bet (as in standard preference reversals), choose the certainty over the P bet, but then choose the $ bet over the certainty. This pattern of non-transitive choices is exactly that which is predicted by regret theory and is analogous to preference reversals because it entails the P bet being chosen over the $ bet. This result obviously challenges any theory that explains choice in terms of transitive preferences.

4.16.2 Psychological explanations for preference reversal

Psychologists believe that preferences are context sensitive. That is, preferences are formed during the performance of particular tasks and there are different preferences for different classes of task (Cubitt, Munro et al. 2000). This means that the process of choosing your preferred category and then attaching a monetary value to it involves two different thought processes. Individuals tend to place high valuations on gambles that generate large prizes, even if the probability of winning is low. This means that gamble Q (in example 4.1) is often valued more highly than gamble R (Heap, Hollis et al. 1992).

4.16.3 Consistency of preferences

Theoretically all respondents have what is known as reference prices (r) for products and when asked if they would be willing to pay a bid amount (b) for that product they will say 'yes' if $b < r$ and 'no' if $b > r$. If respondents answer in this manner then they are being consistent with their preferences. However in a hypothetical framework how does one determine if the individual is responding in an equivalent way to what they would in a real-life framework. The literature on contingent valuation (CV) points to many reasons why subjects may be inconsistent with their real preferences (Basset 1997).

- A misunderstanding of the question or the scenario being presented to the respondent. Quite often in environmental economics, the accidents are 'seldom as bad as we are lead to believe' (Basset 1997). This is a fundamental prerequisite for a CV survey to be valid – the respondent must always understand what it is that they are valuing.

- A CV question may lead to a 'protest' response. The subject may have an aversion to the method of payment due to a moral standing. For example, when being asked to contribute to public services, the respondent may feel that they have already done their fair share by paying their taxes therefore claim that it is up to the government to do the funding. The respondent misunderstands the hypothetical nature of what is being asked of them. In environmental economics the subject may think that it is up to the 'industry' to deal with the problem – they should be held responsible.
- The respondent may fail to consider the substitutes (private or public) available for the good being valued. This will lead to an invalid positive response. Alternatively, the respondent may feel a 'warm glow' from contributing to a service that they feel is benefiting society.
- The subject could be trying to use the survey as a game whereby the answer they give will affect the likelihood of the service being implemented. If the respondent feels that they will actually have to pay for the product if it is introduced then they will revise their bid downwards. Alternatively, if they feel that they would not have to actually pay, but their bid amount will affect the likelihood of the service being introduced; they will revise their bid upwards. By giving a high bid, they think that the service is more likely to be introduced. This is known as strategic bias.
- Depending on how the CV values are elicited, the respondent might feel that they are being asked for the 'right' or the 'good' answer. This is more likely to happen

when the person is being interviewed as they are 'put on the spot' and therefore more likely to be put under pressure to give (what they feel) is the 'right' answer.

- The subject may be revealing a value that represents a value for the whole class of products being evaluated (the embedding effect).

Within a CV survey, all of the above potential biases can lead to people giving answers that are inconsistent with their reference prices and therefore inconsistent with their preferences. A CV survey should therefore be designed to try and achieve responses that are identical to that revealed under a real-life referendum.

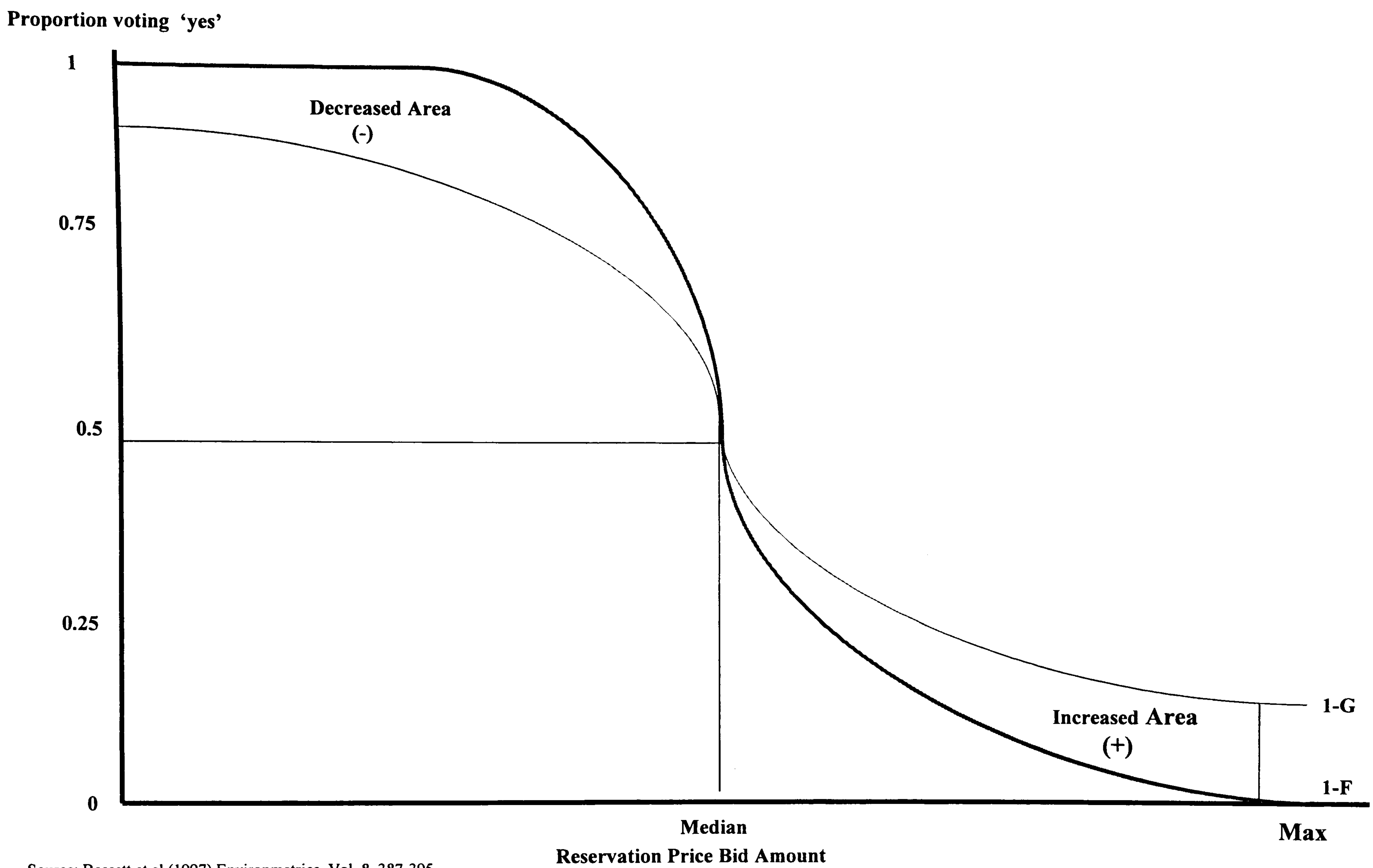
When the respondents depart from what they would do under a real-life setting, they can be referred to as 'yea-sayers' and 'nay-sayers'. Yea-sayers answer 'yes' regardless of the bid amount that they are being presented with. Nay-sayers respond in the opposite way and they view the survey in such a way so that they answer 'no' to every survey question asked. Why is this so? The affect this group can have on the overall survey results can be huge and they present a real problem as to how to deal with them. It is difficult to ascertain what determines a 'yea-saying' response and a 'nay-saying' response. How do we know if that person has given a genuine 'yes' response or are not answering the question truthfully? Also, what lies behind a 'yea-saying' or a 'nay-saying' response – if the researcher knew the answer to this, then they could design the survey so as to try and eliminate them. This is where it is important to mention the role of qualitative work in CV surveys. Asking respondents after they have responded to the CV question for their reasons why they responded in such a way could aid the researcher in identifying any 'yea' or 'nay' sayers. According to Alan Shiell, '*If we are to understand better what people mean by the* *answers that they give to economic surveys then we need to learn from and trust the insights that come from qualitative research techniques*' (Shiell, Seymour et al. 2000). The reasoning behind a CV response can help to determine, among others, the protestors to the survey design, people who have the 'warm glow' feeling, subjects who have attempted to estimate the resource costs of the product and all those that have produced an embedding effect. This qualitative data can then be used to either validate or invalidate the quantitative response.

If a survey design leads to respondents being inconsistent, what affect does 'yea-saying' and 'nay-saying' have on the overall WTP results? As Bassett et al. (1997) predicts, there are too many 'no' votes at low bid amounts, and too many 'yes' votes at high bid values (Basset 1997). The problem however is that these two affects do not cancel each other out. The overall mean WTP value is always biased towards the upward trend because the magnitude of 'yea-sayers' is always greater than 'nay-sayers'. Bassett modeled what happens to the WTP response curve when 'yea-sayers' and 'nay-sayers' are present in the sample. This response curve is illustrated in figure 4.2.

The observed response curve is denoted by 1 – G compared to the true response curve (1-F). The effect of 'nay-saying' on the observed response curve leads to it being initially too flat at the low bid levels - too many respondents are saying 'no' to the low bids. At the other end of the curve, too many subjects are saying 'yes' to the high bids resulting in the observed response function flattening too quickly and not approaching zero at the high bids. As the curves clearer show, this leads to a decrease in consumer surplus at the low bids and an increase at the high bids. As Bassett et al.

points out, when the reference prices in a population are normally distributed the affect of yea-saying and nay-saying in the WTP response curve cancel each other out. However most consumers in the population have low reservation prices compared to a small proportion that will have high reservation prices this therefore leads to a skewed distribution in reservation prices. This leads to the overall estimated mean WTP being biased upward as the area in the upper tail between 1-G and 1-F is much larger at the higher bids than the decreased area at the lower bids.

Figure 4.2 Response Distribution with yea and nay saying.



Source: Bassett et al (1997) Environmetrics, Vol. 8, 387-395.

According to the NOAA panel, 'There is no strategic reason for the respondent to do other than answer truthfully' (Arrow, Solow et al. 1993). The NOAA panel recommends that contingent valuation questions should be administered using a hypothetical referendum format therefore the question would take the form: -

'Would you vote for program A if your cost for program A was £B per year?'

The respondent would answer 'yes' if they were WTP for program A and 'no' if they were not WTP for program A and there would be no strategic reason to answer otherwise. Critics of the hypothetical referendum format however argue that there would be no strategic reason for the respondent to answer truthfully. The number of experimental studies examining preference formation within a hypothetical structure has increased rapidly in recent years however still the debate has not been resolved.

For a preference ordering to exist, an individual's preferences have to be complete (Shiell, Seymour et al. 2000). Completeness suggests that for any two bundle of goods, x_i and x_j, where the utility function is denoted as u(.), either $u(x_i) \succsim u(x_j)$ or $u(x_j) \succsim u(x_i)$ or both (which gives indifference ~). In other words when an individual is faced with two bundles of goods, he or she is always able to say whether he or she prefers either one or is indifferent between them.

However there is doubt upon whether individual preferences are well formed especially when it comes to valuing health states. According to Fischhoff values only come to be recognised when an individual has the opportunity to act upon them and reflect on the consequences of their decisions (Fischhoff, Bostrom et al. 1993). It is probable that individual values may be constructed during the elicitation process, that is, the act of eliciting the values that people attach to health states may in itself prompt a process of 'value clarification'. This has an important implication when measuring the 'performance' of elicitation techniques. Fischhoff argues that the individual may 'try on' a value in order to see how it fits, they then reflect on their answer, and possibly revise their response. The process of elicitation therefore may aid the construction of preferences as well as elicit them and any discrepancies in result may be a function of a deliberate process of reflection and revision as opposed to a measurement error (Shiell, Seymour et al. 2000).

Shiell et al. (2000) constructed an experiment to attempt to test for completeness of preferences. The study aimed to test whether people change their preferences during the process of elicitation and revise them accordingly. The experiment looked at the pattern of responses over time. Using the Standard Gamble (SG) approach, individuals were interviewed and asked to place scores on two health states at three separate points in time. The individuals in the study were also asked at the end of the second and third interview if any significant events had happened in their lives so as to control for the effect this may have on their underlying preference structure. Forty per cent (16 individuals) of the sample in the study attributed stable values to the two health states over time therefore the authors suggest that these preferences may be complete. A sizable proportion of the study sample reflected on the value of health and consequently, after reflection (from the first to the second or third interview), their preferences became more stable over time. The authors do state that it is premature to draw conclusions from such a small and selective sample however there is evidence that the assumption of completeness cannot be taken for granted (Shiell, Seymour et al. 2000). Oliver however challenged some of the conclusions drawn from this study (Oliver 2000). Oliver suggests that the greater stability of values over time is more likely due to a learning effect of the SG approach rather than completeness. Oliver comments that completeness is difficult to test for as subjects may indicate a preference for a particular health state through sheer embarrassment at appearing foolish rather than due to an underlying preference. Therefore he challenges the claim that Shiell's experiment actually tested for completeness of preferences in the first place. Shiell however is confident that the use of the SG approach to measure preferences did not lead to any substantial problems judged through the verbal comments obtained from the subjects during the elicitation process. The people in the study did not say that they could not understand the method or what was required of them. Shiell also points out that the paper did not state that preferences are complete rather they suggested that completeness could not be assumed (Shiell, Hawe et al. 2000).

In a more recent study conducted by Shiell and Gold (Shiell and Gold 2003) the assumption of 'completeness' was examined using a mixed-methods approach, i.e. a questionnaire and interview study. Using the payment scale format, 105 participants completed a questionnaire indicating their WTP for a vaccination programme within the range from zero to £40. Having completed this WTP exercise, the participants were then given a brief structured interview to ascertain the reasons why they had left unmarked values in the range and to establish whether any unmarked values above the stated maximum WTP meant they might pay more than this. Respondents were also informed of the actual cost of producing the composite vaccine (£Y) where this amount was set at £10 more than the minimum amount that the respondent had previously said he or she would most definitely *not* pay – they were then asked if they would WTP this amount. The study found that more individuals left unmarked values above their stated maximum than below it; that the payment scale instructions caused difficulty for some individuals who found it difficult to distinguish between *definitely* would pay and the maximum amount they *thought* they would pay; many respondents indicated that the range of unmarked values above their maximum amount meant that they would pay more than this for the vaccine if saw reason to; nearly one-third of the sample said that they would certainly pay £10 more than the maximum amount previously stated. What the outcome of this study shows is that it is relatively easy to get individuals to increase their stated WTP above their stated maximum by introducing very little extra information. It is possible that the range incorporated by the payment scale may have caused an embedding effect but to minimise this the study only analysed individuals who indicated a point on the scale that they would definitely *not* pay. It is also possible that the design of the study caused interview bias however the sample of respondents used in the study comprised health service researchers who were senior to the interviewer and who are aware of the importance of not being affected by this influence.

4.17 NOAA

Sections 4.2 – 4.15 of the discussion has focused on the debate within the literature concerning aspects to do with the appropriate design of CV studies. Most of the recommendations made, particularly in health care are conjectural. They are based on results of studies, none of it is set in stone and the debate of what is right or wrong is still very much continuing. However, within environmental economics, in 1993, the National Oceanic Atmospheric Association (NOAA) published a set of guidelines advising how CV studies should be conducted. These guidelines first of all criticise aspects of the CV methodology followed by recommendations on design issues. An outline of the debate on CV within health care would not be complete without consideration of this set of guidelines. The focus of recommendations and guidance provided by NOAA are on assessing environmental damages however the guidelines have been cited frequently in the health economics literature. The next section of the chapter summarises the report produced by NOAA.

In response to NOAA recognising that there are 'non-use' or 'passive use' values that have to be measured when assessing environmental damages, an expert panel was requested to evaluate the use of the contingent valuation method. Evaluation involved the expert panel receiving hundreds of pages of comment and holding a public meeting to hear all sides of the debate. Once this process was complete and the panel had evaluated all the comments, a report was issued to NOAA. To summarise the report, first, the criticisms of the contingent valuation method are outlined and second the recommendations from the expert panel are discussed (Arrow, Solow et al. 1993).

- Criticisms of CVM

To begin, the board recognised that self-reported WTP values are significantly higher than actual WTP values. Given that the scenario is hypothetical, stated WTP values are always greater than actual WTP values. The board reported that results from CVM studies appear to be inconsistent with basic assumptions about rational behaviour, i.e. a concern that WTP does not increase with the size of the good. The board felt that without objective standards in which to judge WTP values it is difficult to place confidence on the results of the CVM studies. The difficulty in separating the

WTP for individual environmental issues to the WTP for global environmental issues was also seen as a problem. For example, if you asked an individual for their WTP to clean up an oil spill they are likely to express a value without taking into consideration other programmes that they might want to contribute to therefore the resulting WTP value seems unrealistically large. The board also felt that there had been very little application of the CVM technique that reminded the respondents of any budget constraints. This means that when respondents express their maximum WTP values, they do not consider how much disposable income they have. Within a CV study, it is difficult to verify that the respondent has fully understood what it is that is being valued. If respondents are not provided with certain types of information or they misunderstand the scenario for which they are valuing then it is hard to establish this from the expressed WTP value. The board felt that the values given reflect feelings of public spiritedness and the WTP value is not the respondents actual WTP as it is a value given because the respondent feels good from contributing something to society, i.e. the 'warm glow' effect. Respondents may 'protest' and give an understated WTP for several proposed reasons. The respondent may feel that the programme is never going to be initiated anyway so there is no point in expressing a positive WTP value, they may have difficulty in accepting the hypothetical scenario due to aversion in paying taxes or a view that someone else should subsidise any improvements. The board also expressed concern about the potential of an embedding effect within CV studies and viewed it as a potentially serious problem.

- Recommendations in design

NOAA advises that the use of the open-ended question format, when eliciting WTP values is unlikely to provide the most reliable valuations. This is because in everyday life, respondents are rarely required to place a monetary value on a public good, the scenario therefore lacks realism. This means that the respondent is likely to concentrate on parts of the scenario that are trivial. Also, since there is no guidance with an open-ended request, the respondent has an increased incentive to strategically overstate their WTP value. The panel reported that the use of a range of values (payment scale) from which the respondent is to choose from creates anchoring and other forms of bias. It is recommended therefore to use a dichotomous choice question as this type of design resembles real-life decisions that respondents make everyday. This is because respondents are asked to vote for or against a particular level of taxation which is not uncommon with what happens in real referenda. The advise is to ask the dichotomous choice question in an interview situation if possible as the panel believes it unlikely that reliable estimates will be obtained from mail surveys. There is a recognition for the potential bias within interviews and the panel suggests that surveys should test for interviewer effects, i.e. by getting the respondents to write down their chosen values on a piece of paper. NOAA suggests that the research question should be designed to test for the scope effect. Theoretically, it is expected that the WTP for any good should increase with the number of items being valued therefore the study should be designed to measure this difference in WTP. The actual design of the CV question should provide respondents with a 'no answer' option as well as a 'I would not contribute' option. It is also suggested that follow-up questions be included to find out the characteristics of respondents who are not willing to contribute. It is recommended that all CV reports should make clear the objective of the study, the population sampled, the sampling frame used, sample size and the overall sample non-response rate. The panel also recommended that the WTP question should be used over the WTA question as that would lead to more conservative estimates that are more reliable.

4.18 Discussion

Hanemann believes that some of the influences and biases discussed in this chapter are functions of surveys in general and are not just related to contingent valuation surveys. As Hanemann states;

One cannot avoid the fact that surveys, like all communication, are sensitive to nuance and context and are bound by constraints of human cognition. One tries to detect discrepancies and repair them, but they cannot be entirely ruled out. It is important to keep a sense of proportion. As far as I know, nobody has stopped using data from the Current Population Survey, Consumer Expenditure Survey, Monthly Labour Survey, or Panal Study on Income Dynamics because there are response effects in such surveys. The same should apply to contingent valuation surveys (Hanemann 1994).

In this chapter we have discussed a number of issues with regard to the methodological qualities of the CV method. Although as Hanemann states some of the influences and biases are functions of surveys in general, many of these issues are particular to CV studies and require future research to test and refine the method. Through conducting the literature review for this chapter, a number of areas for future research have been identified and are summarised in table 4.2.

Table 4.2 covers a wide range of methodological components of the CV technique and it would be beyond the scope of this thesis to deal with all of these. The main objective however is to research 'point 1', primarily, the appropriate choice of elicitation format. To do this, four empirical CV studies have been designed to explicitly deal with this issue, and each will be presented in later chapters. The primary concern is to test for the 'performance' of each elicitation format and in so doing a number of methodological issues are dealt with that are presented in table 4.3. In particular the anchoring effect, range bias, choice of instrumentation technique, embedding effect, consistency, the warm glow effect and strategic bias are all discussed at various stages when comparisons are made between the elicitation formats. Table 4.3 provides a summary of the different methodological areas that are dealt with in the proceeding chapters presented in the thesis.

Table 4.2 Areas for future research

Methodological components	Description
1. Choice of elicitation format	- Open-ended - Payment scale - Iterative bidding - Closed-ended - Closed-ended with follow up - Contingent ranking
2. Scope effect	To check that WTP is sensitive to the size of the good, the levels of the good should be varied to assess if the WTP values move in the direction expected.
3. Convergent validity	Values elicited through the CV method can be compared to alternative outcome measures such as TTO, SG to check for the convergent validity.
4. Criterion validity	Comparison of stated WTP values with revealed WTP values measured through an alternative method such as travel cost method.
5. Ordering effect	The WTP results, in theory, should not be sensitive to the ordering of questions.
6. Anchoring effect	- Starting point bias: the final WTP should not be influenced by the starting value in a bidding process. - 'Yea-saying': respondents should not say 'yes' regardless of the bid amount.
7. Range bias	Influence on the WTP values by range of values inserted into the payment scale design.
8. Choice of instrumentation technique	- Interview (personal or telephone). - Mail questionnaire - Focus group discussion
9. Embedding effect	This happens when there is a tendency to consider the global good being valued instead of the individual alternatives.
10. 'Warm glow' effect	Individuals may state any WTP value due to the moral satisfaction element of WTP.
11. Strategic bias	Deliberate over/understatement of WTP value due to perceived influence on policy.
12. Consistency	Consistency between direction of preference and WTP values.

Table 4.3 Research issues dealt with in thesis

Chapter	Research element
Six	- Comparison of open-ended with payment scale elicitation format. - Response rate. - Consistency. - Interpretation of zero values. - Embedding effect. - Criterion validity (comparison of WTP with travel cost estimates). - Use of mail questionnaire. - Comparison with qualitative data revealed.
Seven	- Comparison of two lengths of scale used in payment scale elicitation format. - Investigation of range bias. - Use of mail questionnaire
Eight	- Investigation of the closed-ended elicitation format. - Use of mail questionnaire. - 'Yea-saying' effect. - Response rate. - Consistency - Comparison with qualitative information revealed.
Nine	- Use of the iterative bidding elicitation format. - Use of personal interviews. - 'Starting point' bias. - Response rate. - Consistency. - Comparison with qualitative information collected. - 'Opportunity cost' investigation. - Warm glow effect
Ten	- Overall comparison of elicitation formats.

The literature review presented in this chapter has been set out to provide a base from which all further discussion in the thesis will stem from. Any papers that have been not cited within the main text of the chapter have been summarised in the table presented in the appendix. Rather than deal with all of the issues presented in table 4.2, the thesis will investigate some in depth with the intention that the results (and conclusions) will be a substantial contribution to the development and refinement of the application of the CV method in health care.

Throughout the empirical research presented in the thesis, the application of the CV method is on measuring the value of colorectal cancer screening. Before moving to the empirical section the next chapter provides a brief summary of the background to colorectal cancer screening and screening in general. Previous work on using the CV method to estimate preferences for screening programmes is also discussed.

Chapter Five – Cancer and Screening

5.1 Introduction

This chapter is designed to give background knowledge to the disease area discussed in the empirical section of the thesis. Although the emphasis of the research is on investigating the methodology of the contingent valuation method, in the studies presented in this thesis, it is applied to value colorectal cancer screening. Therefore it is useful to provide a brief outline of the nature of the disease and research that has been conducted to examine the cost-effectiveness of the alternative screening programmes.

5.2 Cancer

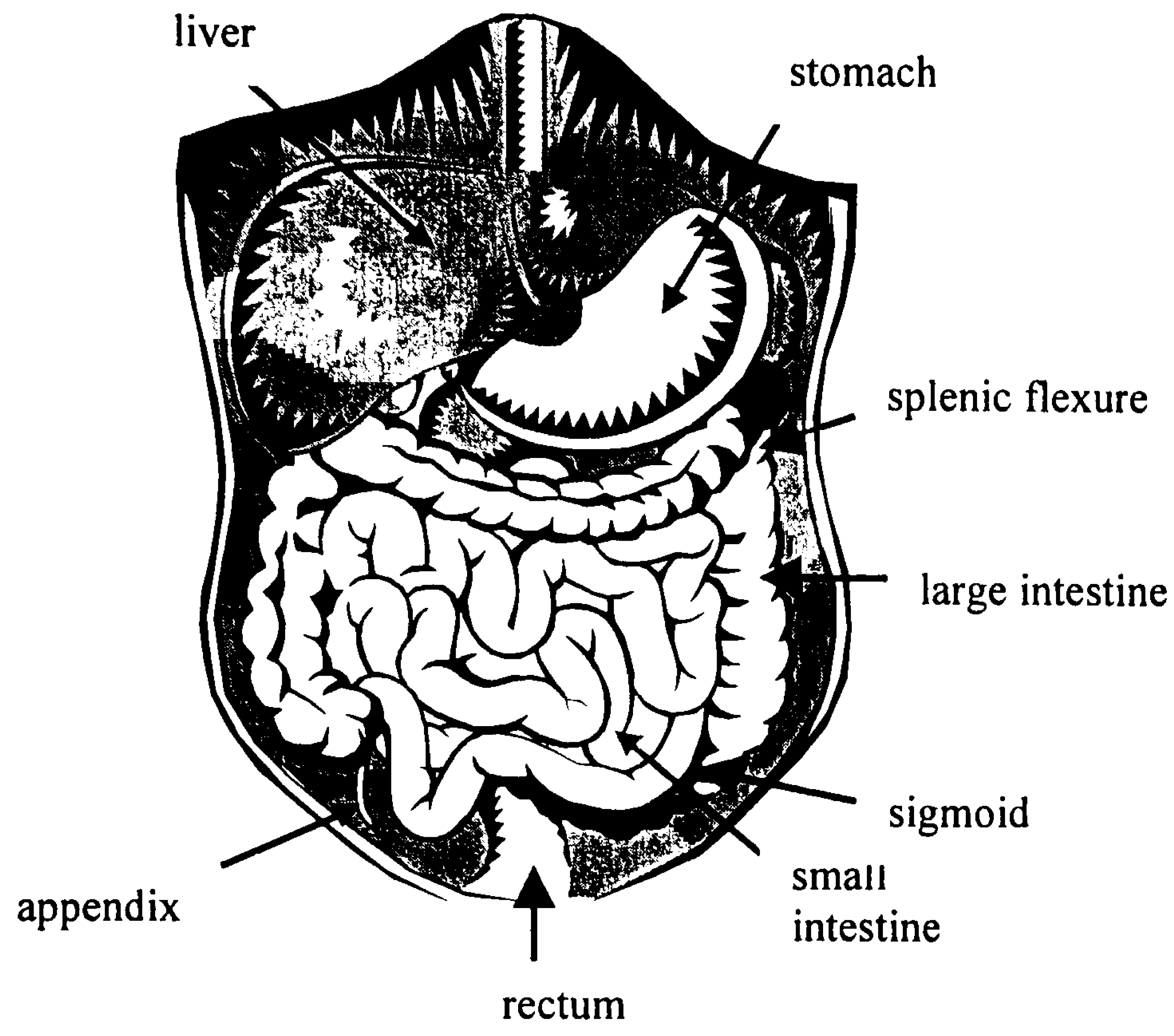
Cancer is a disease that is caused by a malignant abnormal growth of tissue (tumour). It arises from the abnormal and uncontrolled division of cells that then invade and destroy the surrounding tissues. Each individual primary tumour has its own pattern of local behaviour and metastasis (spread). In the European Union, there are 1,631,932 cases of cancer recorded annually (Cancer 1996).

5.3 Colorectal Cancer

5.3.1 Introduction to the disease

Colorectal cancer is cancer of the colon or rectum. The colon is the main part of the large intestine and has no digestive function but absorbs large amounts of water and electrolytes from the undigested food passed on from the small intestine. The rectum is the terminal part of the large intestine that stores faeces before defecation.

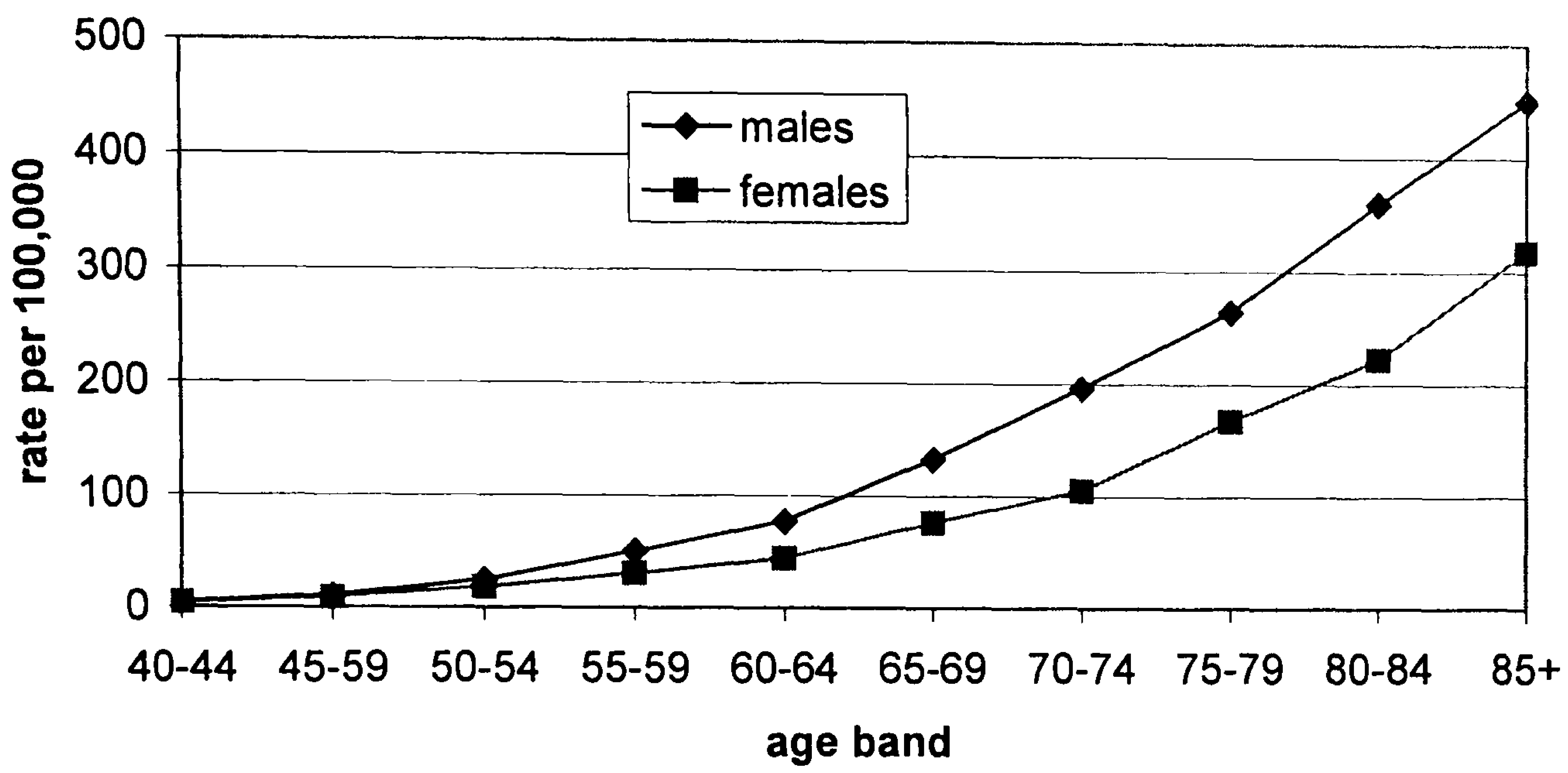
Figure 5.1 The colon and rectum



Colorectal cancer is essentially a disease of developed countries with the highest rates of incidence in North America, Northern and Western Europe and New Zealand. The lowest rates are found in Africa and Asia (OPCS 1994). In the European Union, in 1996, there were 220,973 cases of colorectal cancer (13% of all malignancies) (World Health Organisation 1997). Colorectal cancer is responsible for over 110,000 deaths in the European Union (11% of all cancer deaths) (EUCAN 1996).

The majority of patients that develop colorectal cancer have no predisposing factors for the disease. Factors that have been associated with an increased risk for the development of the disease are older age, family history of colorectal cancer, certain hereditary conditions, inflammatory bowel disease, diets that are high in saturated fat and/or low in fibre, excessive alcohol consumption and a sedentary lifestyle. Figure 5.2 illustrates the age-standardised rates per 100,000 individuals.

Figure 5.2 Age standardised rate per 100,000 (1998 females, 1999 males) (WHO 1997)



5.3.2 Adenoma-carcinoma sequence

It is generally accepted that colorectal cancer begins in the colon or rectum as a cluster of abnormal cells known as adenomatous polyps. It therefore follows that if adenomatous polyps are identified in the colon or rectum they are normally removed to eliminate the risk of developing cancer. This means there is very little direct evidence to support the adenoma-carcinoma sequence. However examination of the polyps show that they have the same anatomic distribution as cancers. Patients who have polyps that are 1cm or greater are found to be at an increased risk of developing future cancer (Winawer, Zauber et al. 1987), the majority of these cancers develop at the same site as the large polyps if left in place (Stryker, Wolff et al. 1987). It is estimated that 1cm polyps take approximately 10 years to progress to invasive cancer (Selby, Friedman et al. 1992). The overall size and type of polyp determines the risk of developing cancer with polyps smaller than 1cm associated with a 1% risk, polyps between 1 and 2cm, a 10% risk, and polyps larger than 2cm, a risk of more than 25% (Muto, Bussey et al. 1975).

5.3.3 Lifetime projections

In the late 1980s, in the United States, estimates of the lifetime probability of developing colorectal cancer were calculated to be 6.0% for women and 6.2% for men. These estimates were based upon life table projections of current annual age, sex, site-specific incidence rates and consideration of competing causes of death that remove individuals from the subsequent risks of developing cancer.

5.4 Staging colorectal cancer

The Duke's system is one of the oldest and most commonly employed colorectal cancer staging systems. In 1930, Cuthbert Dukes developed a three-letter classification system for rectal cancers which was later revised by Turnball in 1967 by the inclusion of a 4th stage, (table 5.1).

Table 5.1 Duke's staging system

STAGE	DESCRIPTION
A	Indicates the least severe disease state: the cancer penetrates into but not through the bowel wall.
B	Represents penetration through the bowel wall, but no invasion of the lymph nodes.
C	Indicates involvement of the lymph nodes regardless of the extent of bowel wall penetration.
D	The most advanced stage, indicates the presence of a primary tumour, lymph node invasion, and the presence of distant metastases.

The tumours typically progress from normal mucosa, through adenomas/polyps, to invasive carcinomas with increasing spread. The prognosis is strongly correlated with the rate of progression at time of diagnosis and treatment. Patients whose cancers are detected at an earlier stage (Duke's stage A and B) have an 85% 5-year relative survival rate compared to 38% in patients with late cancer (Duke's stage C and D), (U.S. Dept of Health and Human Services, 1989).

5.5 Treatment of colorectal cancer

Once the cancer has been detected, there are a range of options for treatment according to the type of cancer and the health status of the patient. Surgery remains the accepted form of treatment for colorectal cancer and is judged to be potentially curative as there is no evidence of residual disease. However when the cancer is advanced, adjuvant radiotherapy is beneficial for palliation purposes. The use of radiotherapy helps to alleviate distressing symptoms of the condition such as pain. There is evidence to show that radiotherapy leads to a reduction in local recurrence rates, (Gerard, Buyse et al. 1988). Chemotherapy might be used to provide a marginal improvement in disease free intervals and mortality (Gastrointestinal Tumour Study Group 1984). And finally, hormone therapy might be used but even though antagonists and antibodies have achieved some success in controlling cancer growth, more work is needed before the role of hormone therapy is properly assessed (Hoosein, Kiener et al. 1988).

5.6 The rationale for screening

"Screening refers to the application of a test to people who are as yet asymptomatic for the purpose of classifying them with respect to their likelihood of having a particular disease."

(Hennekens and Buring 1987).

The implicit assumption underlying the concept of screening is that patients whose cancers are detected in earlier stages will have a more favourable prognosis than patients with more advanced disease on detection. The screening procedure will not lead to an immediate diagnosis. Patients who test positive will be sent on for further evaluation by a subsequent diagnostic test to determine whether they do in fact have the disease.

The value of a screening test is never self-evident. There are often risks or costs involved in the screening procedure and/or consequent diagnostic procedures that must be weighed against the benefits. For a disease to be amenable for screening, one question of importance is whether treatment of preclinical disease is more effective than treatment begun after the development of symptoms. To address this question the natural history of the disease has to be considered.

In 1966, the World Health Organisation (WHO) put together a set of guidelines addressing the issues that should be considered before implementing a screening programme for a disease (WHO 1997). These guidelines were later revised by the National Screening Committee (NSC) to take into account the greater concern about the adverse effects of screening; regrettably some people who undergo screening will suffer adverse effects without receiving benefit from the programme. These guidelines have been summarised in figure 5.3, for a full description of the criteria set out to appraise a screening programme the reader is referred to appendix two.

Figure 5.3 The NSC criteria that should be met before screening for a condition is initiated:

The Condition
- The condition should be an important health problem.

The Test
- There should be a simple, safe, precise and validated screening test.

The Treatment
- There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.

The Screening Programme
- There must be evidence from high quality Randomised Controlled Trials that the screening programme is effective in reducing mortality or morbidity. Where screening is aimed solely at providing information to allow the person being screening to make an 'informed choice' (e.g. Down's syndrome, cystic fibrosis carrier screening), there must be evidence from high quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.

5.7 Measuring preferences for screening

Measuring patient preferences for screening has become a popular exercise for health care researchers over the years. This is due to emerging techniques, like contingent valuation, but also because of the increased significance on incorporating patient preferences into the decision making process. Investigating the preferences for procedures like screening programmes can provide valuable information to decision-makers about the likely future compliance amongst different patient groups. This can help in designing the programme to try and achieve a high compliance by satisfying these different preferences.

Gyrd-Hansen and Sogaard investigated the preferences for screening for breast cancer and colorectal cancer in Denmark (Gyrd-Hansen and Sogaard 2001). The authors argue that *'more knowledge of the nature of public preferences for screening programmes and their attributes may contribute to the evaluation of such health care programmes'*. The study looked at whether the different processes of care as a result of the different screening programmes results in different preferences for the cancer screening programmes. Each respondent was given a description of three alternative screening set-ups and the consequences of not entering a screening programme and asked to rank the programmes in order of preference. The respondents were then asked for their reasons for choosing to participate by checking off a list of possible motivations and asked which of these factors had the greatest influence on their decisions. The authors were then able to establish the main motivations for participating in colorectal cancer screening programmes and compare them to reasons for participating in breast cancer screening. Preferences were mainly explained by disutility associated with out-of-pocket expenses and the utility associated with mortality risk reduction, the receiving of information and the elimination of regret. Anxiety and unpleasantness associated with the screening tests as well as the high risk of a false positive diagnosis were also given as additional reasons not to participate in the programmes. The reasons for participation in the breast cancer screening programmes coincided with reasons for participating in the colorectal cancer screening programmes. The results also implied that the public had consistent preference structures that were generisable so that significant parameters were identified.

Donaldson et al (1997) investigated the preferences for two methods of screening for cystic fibrosis carrier status. Questionnaires were administered to women attending antenatal clinics at Aberdeen Maternity Hospital. Both stepwise and couple screening methods were described and the women were asked which of the two methods they preferred. The results reported that 61% of those who responded had a preference for the stepwise screening method whilst 27% had a preference for the couple screening method (12% had no preference). These results suggest that, in theory, since the majority of respondents prefer the stepwise method then it should be the stepwise method that is provided (Donaldson, Shackley et al. 1997).

The benefits of screening for prostate cancer are uncertain. Volk et al (1997) measured the preferences of husbands and wives for prostate cancer screening. This study used a decision analytic approach to explore the preferences of 10 couples and found that 7 out of 10 husbands preferred the no screening strategy, while 9 out of 10 wives preferred screening for their husband. This was because the wives attached little burden to the complications of treatment, choosing to maximise their husband's life whereas the husbands were more concerned about these complications. The implication of this result is that optimal screening strategies may differ for husbands and wives and guidelines for screening should consider assessing preferences on an individual couple basis (Volk, Cantor et al. 1997).

Eaker et al (2001) set out to investigate how attitudes and beliefs about Pap smear screening affect women's choice to participate in organised or opportunistic cervical screening. This study used telephone interviews on 430 non-attenders and 514 attenders to Pap smear screening in the Uppsala County, Sweden. The study found that important differences in attitudes and beliefs about the test appear between the non-attenders and the attenders. Non-attendance was negatively associated with perceived severity of cervical cancer compared to other malignancies, as well as with satisfactory benefits, but positively associated with time-consuming and economical barriers. The authors state in the conclusion that rather than being emotional, the main barriers are either practical or rooted in misunderstandings and lack of relevant information related to the screening. These insights therefore offer opportunities to increase attendance rates considerably (Eaker, Adami et al. 2001).

In the US, a survey consisting of 17 questions assessing satisfaction and demographics was distributed to women whose screening mammograms were either read 'online' (results are communicated immediately), or 'offline' (mammograms are read in batches the next day and the results are communicated at a later time). Two samples of women (split randomly between the online and offline method) were analysed. The study found the overall mean satisfaction with the mammographic experience and with the time taken to receive results differed significantly between the two groups. Results showed that if the women were allowed to choose between the two methods, 97% of patients in the online group and 91% of patients in the offline group would choose online interpretation (Wilson, Wallace et al. 1998).

Serlin et al (2002), assess the preferences of sexually active adolescents towards 3 screening methods for detection of Chlamydia trachomatis, Neisseria gonorrhoeae, and Trichomonas viaginalis using first-void urine (FVU), self-collected vaginal swab specimens, and pelvic examination with clinician-collected endocervical swab specimen collection. The adolescents were asked to rank the 3 screening techniques according to preference. The adolescents preferred the FVU for sexually transmitted disease screening over the pelvic examination and the self-administered vaginal swab test. They viewed the FVU the most positively, the pelvic examination the most negatively and the vaginal swab technique slightly less positively than the FVU (Serlin, Shafer et al. 2002).

The above studies that have measured the preferences of individuals for different screening programmes are just meant as a summary of the huge amount of research that has been conducted in this area. It seems that incorporating patient preferences is important with any health care programme but it can be argued more so with screening programmes as the preferences can directly affect the future compliance. Screening is unique in that the patients are not directly suffering from illness when asked to come forward for a test, a greater responsibility therefore lies with the patient to understand the importance and the reasons why they are being screened. With the knowledge of the different preferences across patient groups, decision-makers are better placed to 'design' programmes to achieve maximum patient satisfaction and improve compliance.

5.8 Methods of screening for colorectal cancer

The primary objective of colorectal cancer screening is to detect the disease at its pre-clinical stage, i.e. the detection of adenomas, polyps. There are a number of tests available (Young, Macrae et al. 1996) that have been designed to detect pre-clinical colorectal cancer and each will be discussed in turn.

1. Digital Rectal examination
2. Barium enema examination
3. Rigid sigmoidoscopy
4. Flexible sigmoidoscopy
5. Colonoscopy
6. Occult Blood Tests

5.7.1 Digital rectal examination

This procedure involves the clinician using a finger to examine the rectum and anal region for any abnormal tissue. This method, although useful and relatively straightforward as a preliminary analysis, is limited as only concentrates on the rectum and not the colon.

5.7.2 Barium enema examination

Before the introduction of fibreoptic endoscopes the principle method for detecting colorectal cancer was to use the barium enema examination method. This method involves using barium to outline the bowel. Liquid barium and air are introduced into the bowel through a small tube. The barium highlights the bowel and X-rays show any irregularity in the bowel wall that may be caused by cancer. To obtain a good result however, adequate bowel preparation is vital as otherwise faecal material can lead to a false positive result.

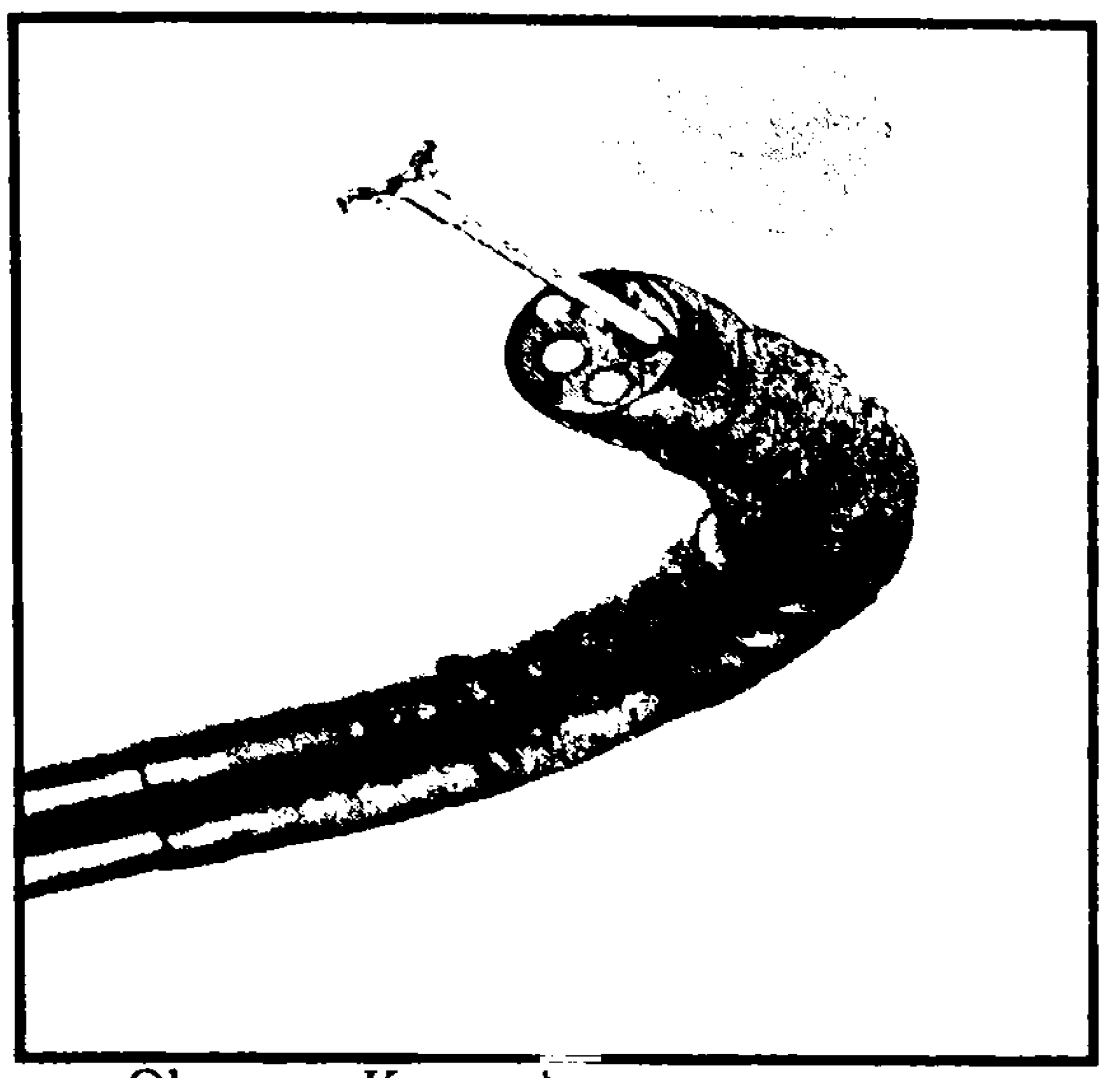
5.7.3 Rigid sigmoidoscopy

The rigid sigmoidoscopy technique has been used as a method of detection since the 1930s. This involves a standard 25cm rigid sigmoidoscope that has an average reach of 18-20cms. A scope is a long thin tube that is an instrument used for examination. Approximately 40-45% of all colorectal cancers should be detected by the instrument, however, in practise the instrument is not inserted fully (Nivatvongs and Fryd 1980) (Wilking, Petrelli et al. 1986). If the bowel is not properly telescoped and proper insertion procedures are followed, the bowel may be stretched to 25cms and there is about 0.01% chance of colonic perforation (Winawer 1980).

5.7.4 Fibreoptic sigmoidoscopy

Examination is conducted using a flexible telescope, which is passed up from the anus. Figure 5.4 illustrates a typical flexible telescope used to visualise the bowel.

Figure 5.4 Fibreoptic sigmoidoscope



Source: Olympus, Keymed.

This procedure requires a simple bowel preparation; it can be used quickly in outpatient departments, without sedation. The procedure allows examination of the rectum and colon to 60cms from the anal margin. Any low-risk adenomas found during the examination can be safely and simply excised using the instrument at that time. Any patients found to have high-risk adenomas would be referred on to colonoscopy. As a result of screening using the 'flexi-scope' the patient would be left with a clean distal (60cms from anal margin) colon. After the procedure the chances of developing any abnormal lesions that would have an affect on life-expectancy are so slim that it is proposed this test is only required once during a life-time. Findings show that 70% of adenomas occurs within the reach of the instrument. In a study using 1,045 adenomas 94% of those polyps containing invasive malignancy were located in the sigmoid or lower colon and within the range of a flexible sigmoidoscope.

5.7.5 Colonoscopy

This technique offers the most reliable examination of the bowel, the procedure is highly sensitive. The drawback being that it is time consuming, expensive and there are small risks involved in the procedure such as colonic perforation or haemorrhage (Williams 1986). This procedure has been ruled out as a mass screening tool due to clinic capacity, cost (unit cost, £187 (Walker, Whynes et al. 1991)) and likely long-run non-compliance due to its invasive nature. It should be reserved as an investigation tool for patients with positive tests from the faecal occult blood test, the clarification of doubtful lesions on barium enema and for the assessment of inflammatory bowel disease (Hardcastle and Balfour 1980).

5.7.6 Faecal occult blood (FOB) tests

Bowel tumours are known to bleed sporadically and to deposit blood in the stools. Although FOB tests vary in their chemical functioning, they all aim to detect the presence of such occult blood and are thus predictors of the presence of abnormalities. The FOB test is normally posted out to patients with detailed instructions enclosed. Although the FOB tests have poorer sensitivity (compared to endoscopic procedures) and thus produce a lower yield, the procedure has potential benefits of offering simplicity, safety and cost effectiveness. If blood is detected in the stool then this is an indication that there is an abnormality and the patient is required to have an investigation.

5.8 Rationale for colorectal cancer screening

Colorectal cancer is a disease amenable to screening as it has a high incidence in society (13% of all malignancies), it presents itself in well-defined stages (Dukes classification) and there is evidence to show that treatment given when the tumour is localised to the bowel wall (stage A and B) results in an improved prognosis.

There are acceptable tests available that can be implemented to enable the detection of the disease at its asymptomatic stage, and when the disease is detected there exists safe and acceptable treatment for the condition, e.g. surgery, radiotherapy, chemotherapy.

Routine screening for colorectal cancer is not available within the public sector in the UK. However over the last 20 years, the results of major clinical trials have been published and the evidence is showing that the screening of colorectal cancer can contribute significantly to mortality reduction.

Most of the primary research to date has concentrated on just two methods of screening, the FOB test and the Flexible Sigmoidoscopy (FS) test. Each is therefore a potential (and rival) candidate for mass population screening programme.

5.9 Evidence on FOB screening

Major randomised controlled trials of FOB screening have been undertaken in the past twenty years (Towler, Irwig et al. 1998). The two longest running European trials, conducted in Denmark and in England (Nottingham), have screened people aged 50-74 years biennally (Hardcastle, Chamberlain et al. 1996; Kronborg, Fenger et al.

1996). Economic evaluations conducted alongside these trials have indicated that, if implemented, their protocols would prove cost-effective in comparison to existing screening programmes (Gyrd-Hansen, Sogaard et al. 1998; Whynes, Neilson et al. 1998). The expected cost of FOB screening using the Nottingham protocol, when translated into current prices, emerges as £7.0 per person screened. The total annual screening costs amounts to approximately £44 million (Whynes, Frew et al. 2002). To assess the impact on the general population pilot screening projects are currently operating in the UK (Brooks 1998).

5.10 Evidence on flexible sigmoidoscopy screening

Experimental results for the use of sigmoidoscopy as the principle screening tool have already been published (Newcomb, Norfleet et al. 1992) and, in the USA, the procedure is reimbursable under Medicare (Lewis and Asch 1999). Once only screening is currently the subject of a major UK multi-centre randomised controlled trial (Atkin, Hart et al. 1998), for the purposes of which the procedure has been termed 'Flexi-scope' (FS). Preliminary results from the UK trial estimate that it would cost £52.5 million per annum to screen for colorectal cancer using this protocol (once-only screening) (Whynes, Frew et al. 2002).

5.11 Discussion

This chapter has briefly summarised the nature of colorectal cancer as a disease and presented a case for why it is an ideal candidate for mass population screening. Judging from the clinical trials, it is evident that there is a strong clinical research interest into the two rival potential screening programmes for colorectal cancer (FOB testing and FS testing). What the empirical work in this thesis plans to address however is if there is a corresponding interest among the potential demanders of the screening tests, i.e. the general population.

Although colorectal cancer screening would be provided to everyone free at the point of use as part of a tax financed NHS system, it is not technically a public good. Formally, a true public good is one that is non-rival in consumption (the consumption of one individual does not reduce the benefits derived by all other individuals), and non-excludable (one consumer cannot be excluded from consumption benefits). A good example of a public good is national defence as all individuals enjoy fully the additional security from the existence of national defence and you cannot exclude any one individual from consuming this benefit. It is plausible that individuals can have the option of consuming colorectal cancer screening by paying for it through a private scheme however it is within a public health care context that the programme will be valued in this thesis. This is why tools such as the contingent valuation method are ideal for this purpose as they are commonly used to value true public goods such as environmental initiatives. The empirical work therefore is seeking to measure the preferences of the potential patients that might be screened for colorectal cancer. Using the WTP method the value that the general population places on the screening options will be estimated. Direct comparison will be made between the values elicited for the series of FOB tests compared to the once-only FS test and these will be analysed across sample sub-groups. Chapter six presents the first study.

Chapter Six – Open-Ended and Payment Scale Study

6.1 Introduction

This chapter describes the first piece of empirical work set up to begin the investigation into the willingness to pay (WTP) methodology. One primary aim of this study was to estimate a WTP value for colorectal cancer screening. In the process of getting to this estimate two WTP elicitation formats (open-ended and payment scale) are compared with reference to many of the methodological components of the technique that have been discussed in chapter four.

6.2 Aim

The study had a number of objectives. First, using the willingness to pay (WTP) method the study set out to measure a monetary estimate of the value of screening for colorectal cancer. Second, given the different processes of care involved with the alternative screening programmes the study aimed to assess whether faecal occult blood (FOB) testing is more or less valued or acceptable than flexible sigmoidoscopy (FS) testing, i.e. a comparison of the actual modalities of screening for colorectal cancer. Third, since there are many elicitation methods to choose from when conducting a WTP project, the study set out to ascertain whether the way in which the question is asked affects the responses given, i.e. a comparison of the different elicitation formats. And finally, to assess the impact of demographic factors on the absolute and relative WTP valuations provided for each of the screening programmes.

6.3 Method

6.3.1 Distribution

The WTP data capture instrument was a questionnaire designed for respondents to complete without supervision. After initial construction, the instrument was piloted on thirty members of the general population that were recruited through a local general practice. Following the results of the pilot study, the instrument was revised then presented for ethical approval and again modified accordingly.

'The Trent Focus Network' of collaborative general practitioners (GPs) represented an ideal opportunity for the distribution of the study questionnaires. The organisation is a Collaborative Research Network (CRN) consisting of a set of primary care practices that are located within the Trent region of East Central England. As part of this network, all the practices are willing to contribute to research with the overall network aim of facilitating high quality research within the region. In total, the CRN consists of 68 practices located across the region that covers both rural (Lincolnshire and Humberside) and urban (Sheffield and Nottingham) areas. At the preliminary stages of the study the network was in its first year of development. It offered the potential of easy access to patients and offered the capacity for general practitioners to filter out unsuitable subjects (subjects that did not meet the inclusion criteria for the study). Also, since most subjects drawn from the Focus practices had no experience of either screening modality (FOB or FS), the network offered a sample of individuals that were generalisable to the general population.

To begin, a study protocol was submitted to the Trent Focus CRN Board, the board then assessed the protocol according to pre-defined criteria and accepted it as being a study design that was suitable for the network. The protocol was then posted out to all the CRN practices and if the practices were interested in the study then they offered to participate. In total, twenty-two practices agreed to participate in the study which involved distributing the copies of the questionnaire to their patients on our behalf. Postcode mapping of the 22 practices confirmed that they were well spaced out across both the rural and urban areas. The practices were also not statistically different from the rest of the CRN practices in terms of list size (P = 0.371) and number of partners (P=0.149), using the Mann-Whitney U-test.

In the first month of the study, each of the twenty-two practices received a personal visit to explain the study protocol. After arranging the appointments by phone, I met with the practice manager and usually at least one of the General Practitioners (GPs) from the surgery. These visits to the practice provided an opportunity to 'get to know' the practice staff and to clear up any queries regarding the study protocol. At the meeting each practice were given the questionnaires that were to be distributed. Each GP received eighty-four questionnaires for distribution and since the number of GPs across the twenty-two practices amounted to eighty-three, this meant that 6,972 questionnaires were allocated for distribution. The GPs were requested to offer the questionnaire to any patient during a normal consultation, subject to three forms of exclusion. These were, first, persons under 25 years of age, on the grounds of likely perceived irrelevance of screening in that age group. Second, GPs were at their discretion, to exclude any subject with a recent diagnosis of colorectal cancer in the family, on the grounds of minimising distress. Finally, potential subjects with substantial reading, learning or language difficulties were to be excluded, on the grounds of incapacity to complete the questionnaire. The practices were encouraged to collect in all completed questionnaires and return them to the University of Nottingham in batches of 12-15 in large envelopes provided. Each practice was then reimbursed for all postage cost incurred during the study. If the patient had a strong desire to take the questionnaire home with them to complete, individual pre-paid envelopes were provided for this purpose and GPs were asked to give these out accordingly. Completion at home however was not encouraged, due to expected low completion rates with this method. The practices were given a six-month time period to hand out all the questionnaires. It was at the initial meeting with the practice that I asked the practice manager to allocate one member of surgery staff to co-ordinate the distribution and return of the completed questionnaires. This person then acted as the 'point of contact' and over the six-month distribution period received at least one phone call per month to check on the progress of the study and to clear up any problems.

6.3.2 Data Capture Instrument

The questionnaires were designed as eight-page (double-sided) booklets. The first two pages of the instrument contained standard administration details, such as contact information and a guide to completion, descriptions of colorectal cancer, the principle of screening and the two screening options, FOB and FS. The descriptions used were essentially similar to those which had been employed in inviting subjects to participate in the Nottingham-based English FOB trial (Hardcastle, Chamberlain et al. 1996) and the UK multi-centre FS trial (Atkin, Hart et al. 1998). At the time the study was administered, rates of effectiveness such as sensitivity and specificity levels for each test was not known. The study therefore kept the descriptions of each screening modality to a description of the process that the individual would have to go through when receiving a colorectal cancer screening test. This would be the descriptions that individuals would receive if they were to be invited to either of these screening tests as part of a National Screening Programme, the results gained from this study therefore would help to predict future compliance of either of these screening protocols. The descriptions were kept as simplified as possible to minimise cognitive overload.

Once the subject had read through these descriptions they were then asked whether they would wish to undertake a screening test were one to become available, and if so, which test would they prefer. Thereafter, each subject was invited to supply a WTP valuation for each of the two screening options, FOB and FS.

Two elicitation formats, either the open-ended or payment scale design, were randomly offered to subjects (each GP received a shuffled pile of questionnaires containing equal numbers of each design). The decision to adopt the open-ended and payment scale format was due to a number of factors. At the time the study was designed, open-ended and payment scale formats had gained popularity in the literature. Advice was sought from contingent valuation experts and the open-ended and payment scale designs were recommended due to the ease of administration. Both the open-ended and payment scale designs can easily be described in a postal questionnaire.

The open-ended format asked subjects what their maximum WTP would be for first, the complete series of FOB tests every 2 years and second, the one-off Flexi-scope test at age 60. In the payment scale format, the space provided in the open-ended variant was replaced with a vertically arranged list of 29 values. These values were, from top to bottom, £0, £5 and £10, thereafter to £100 in units of £10, to £200 in units of £20, to £500 in units of £50 and to £1000 in units of £100. Subjects expressing a valuation in excess of £1000 were requested to write in the appropriate amount (i.e. beyond the limit of the scale, the payment format defaulted to the open-ended format). Subjects answering the payment scale format were requested to circle the maximum amount, whilst placing ticks against amounts they were sure they would pay and crosses against amounts they were sure they would not. The structure of the payment scale was chosen as one that had a similar structure to a similar application in the literature (Donaldson and Shackley, 1997)) and on the basis of personal communication from a colleague who had published widely in the area of CV research [C Donaldson]. The open-ended and payment scale questionnaires are presented in appendix three and four respectively.

After the WTP question, the questionnaire then became identical for both formats. Once the respondent had revealed a WTP value they were asked an open-ended question requesting reasons for their answers to the WTP question. The remainder of the questionnaire requested information on a range of socio-demographic and economic variables. Subjects were asked about gender, age, age on leaving formal education, employment status and household income (in four bands, starting at zero, band-width £10,000, ending £30,000 and above). Information was also requested on whether any of the six diseases (stomach problems, haemorrhoids, heart disease, cancer, stroke and depression) had been experienced by the subject or the immediate family, and on perceived own-health status (four point scale, poor, fair, good, excellent). The number of visits to the GP and dentist in the past 1 and 2 years, respectively, was requested, as was any screening history of any type over the previous 5 years (including screening for colorectal cancer). Subjects were asked to note if they were particularly worried about colorectal cancer (four point scale – not at all, a bit, quite, very) and to identify their perceived chances of eventually suffering the condition, compared with men and women of their own age (five point scale – much lower, lower, same, higher, much higher). Finally, the questionnaire requested an indication of the importance they attached to a fruit-rich diet, regular exercise, breast screening and cervical screening in maintaining good health (five-point scale in each of the four cases – not at all, somewhat, moderately, very, extremely). These data were coded 1 through to 5, respectively, and a mean score across the four dimensions were calculated. This mean was then taken as a measure of the subject's orientation towards health promotion or 'health motivation' (Vernon 1997) i.e. the degree of belief in the efficacy of established health promotion measures. Clearly, a higher mean value implies a higher health motivation.

6.3.3 Analysis

The presentation of the results are split into two parts. Part A presents the WTP values that were elicited for FOB and FS screening, part B expands this analysis by examining the methodological aspects of the WTP technique; in particular, looking at the performance of the open-ended versus the payment scale elicitation format.

In part A, the data is analysed using both logistic and linear regression techniques. There are three classes of response in total, no WTP value offered (no response), zero WTP values and positive WTP values. Two separate models are constructed, to explain zero and positive values respectively. Although it would have been feasible to estimate a single Tobit model for the full sample, earlier researchers have found that this two-part specification performs better with WTP data (Donaldson, Jones et al. 1998). All other independent variables were entered into the model and either chosen or rejected according to a backward stepwise selection. The final models are discussed in part A. When conducting the analysis, several of the variables had to be re-coded as binary dummies, for example, a dummy variable for the age data was constructed with the prior hypothesis that subjects under 45 years of age would have a reduced interest in the screening tests given that they are more distant from eligibility for either test.

In part B, the methodology behind the WTP technique is examined in more detail. One such aspect of the methodology is the interpretation of zero valuations as this is unresolved within the CV literature, for example, the distinction between a 'true' and a 'protest' zero. Another methodological issue relates to the formation of preferences. Logic would predict that respondents give a higher WTP value for their preferred test but this result is not always found to be the case in WTP studies. The issue of respondents being 'inconsistent' and the interpretation of zero valuations will be addressed in part B. If it is assumed that protestors and inconsistent respondents have answered the WTP question incorrectly then the WTP values revealed from this group should be excluded from the analysis. Part B will demonstrate what happens to the overall results if this exclusion criterion is adhered to.

The final section of the chapter will examine the policy implications of the study. The estimated WTP values are compared to the actual resource costs of both screening protocols to help to determine whether there is a societal demand for a colorectal cancer screening programme.

6.3.4 Coding of reasons

After answering the WTP questions and before proceeding to the socio-demographic section of the questionnaire, the respondents were asked for explanations as to why they chose their WTP value. The question was worded as a free-text design in which respondents were allowed to write whatever they felt about their WTP.

After reading through the explanations, certain sentiments and phrases appeared regularly, i.e. 'for peace of mind'; 'you can't put a price on health'. To enable an analysis (presented later in the chapter), a coding frame for explanations was developed by iteration. Three members of the health economics group in Nottingham read several hundred responses selected at random, identified themes or categories of explanation, and then endeavoured to position a new sub-sample within this framework. The themes were then repeatedly re-worked between the team members, as and when individual inconsistencies in existing coding appeared to arise.

Part A

6.4 The Results

6.4.1 The Response

A total of 2767 questionnaires were returned, comprising 1366 (49.4%) of the open-ended format and 1401 (50.6%) of the payment scale format. This gave an overall response rate of 40% which was largely influenced by the varying response rates from each of the twenty-two participating practices that ranged from as low as 3% to 96%. The response rate was measured by the number of completed questionnaires returned therefore it is difficult to determine if the rate is affected by the GP failing to distribute the questionnaires or the respondents declining to complete it. The GPs were requested to record 'non-responder' data but unfortunately this data was not collected adequately. For a more detailed discussion on the potential reasons for this wide variation in response rate the reader is referred to (Frew, Hammersley et al. 2001), however to summarise, it is proposed that the number of partners in each practice, structural constraints (building work) and NHS reforms may all have had an influence.

In spite of completing all, or virtually all, other parts of the questionnaires, 553 respondents (20%) failed to complete either of the WTP questions, and a further 293 (10.6%) completed only one. The FOB question received more replies than did the FS question (2156 versus 1979 replies, respectively).

6.4.2 The Sample

Characteristics of the study sample that answered at least one WTP question (n=2214) are illustrated in table 6.1.

For comparison, the UK distribution across the income cohorts is approximately 33, 27, 20 and 20%, respectively (Regional Trends Dataset 1995-1998). This means that the study sample slightly over-represents the middle range of household incomes (£10,000-£30,000), but matches the division between the two highest cohorts combined (≥ £20,000) and the two lowest cohorts combined (< £20,000), almost exactly.

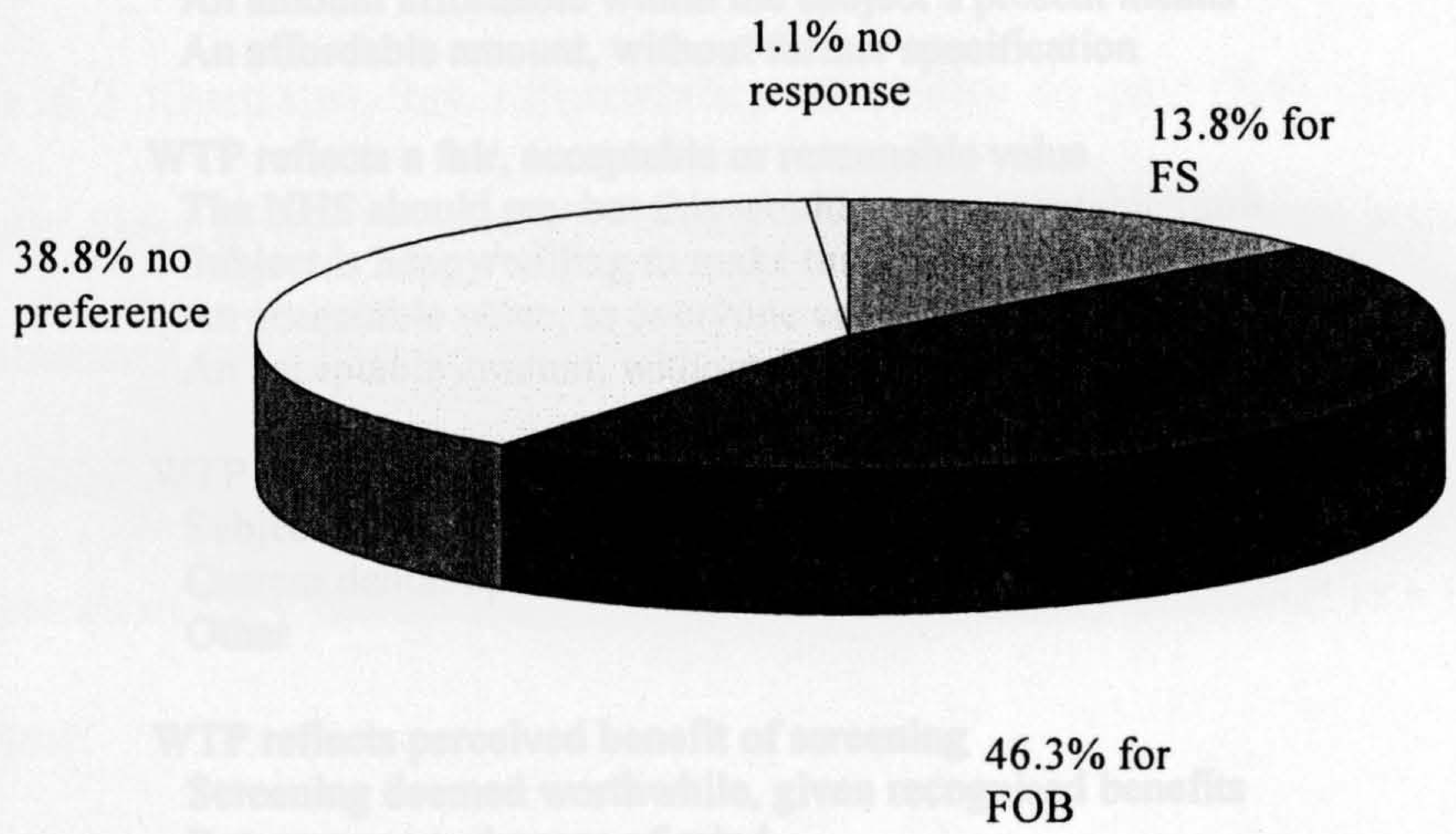
Table 6.1 – Characteristics of the sample

Variables	Sample Characteristics
Gender:	
Female	1720 (62.9 %)
Age [median]	49 years [IQ range 37-60 yrs]
Proportion under 45 years	850 (38.4%)
Income:	
<£10K	550 (24.8%)
£10-20K	760 (34.3%)
£21-30K	494 (22.4%)
≥£30K	410 (18.5%)
Proportion with experience of bowel cancer	188 (8.5%)
Proportion quite/very worried about bowel cancer	456 (20.6%)
% who think chances of developing disease are higher than average	
Number of visits to dentist in last 2 years	309 (11.7%)
Age on leaving full time education [median]	3 [inter-quartile range 1-4]
Smokers (%)	16
Health motivation score [median]	19.6%
	4.25 [inter-quartile range 4-4.5]

6.4.3 Preference groups

Among the respondents that provided at least one WTP, 859 had no preference, 1025 preferred FOB and 305 preferred FS. Twenty-five respondents did not answer this question. The proportions of subjects preferring each test are illustrated in figure 6.1. It is quite clear from the pie chart that the majority of the sample preferred the FOB test.

Figure 6.1: Preference groups



6.4.4 Qualitative Information

As already explained, after answering the WTP question, each respondent was invited to explain their reasons for choosing a maximum WTP value for each test. Of the 1,921 subjects who offered a WTP value for both FOB and FS testing, 1,203 (62.6 per cent) entered a written explanation. Of those offering one WTP only, 54.6 per cent provided an explanation. Before analysing the WTP values, the explanations offered for choosing the values are detailed in table 6.2. The majority of explanations were brief and ranged from a few words to one or two sentences.

Table 6.2: Explanations of WTP valuations

Category	Type of explanation provided	n =	%
E1	Question deemed in-applicable, on the grounds of:	75	3.6
	Subject's age	63	
	Possession of private health insurance	12	
E2	Subject expressed difficulties in estimating WTP owing to:	210	10.1
	Ignorance of cost	30	
	Uncertainty of future financial circumstances	27	
	Difficult to answer	61	
	You can't put a price on health	52	
	Other	40	
E3	WTP estimate started to be based on a nominal amount	64	3.1
	Token or arbitrary sum	51	
	Guess	13	
E4	WTP reflects ability to pay (affordability)	636	30.5
	Stated WTP constrained by pensioners' ability to pay	182	
	Stated value is that which everyone ought to be able to pay	118	
	Maximum affordable, given current income/unemployment	87	
	An amount affordable within the subject's present means	50	
	An affordable amount, without further specification	199	
E5	WTP reflects a fair, acceptable or reasonable value	180	8.6
	The NHS should pay but this would be an acceptable limit	21	
	Subject is happy/willing to make this contribution to NHS costs	21	
	An acceptable value, as everyone could afford this amount	32	
	An acceptable amount, without further specification	106	
E6	WTP reflects costs of screening	215	10.3
	Subject attempted to estimate likely resource costs required	155	
	Current dental/optical charges used as comparator	45	
	Other	15	
E7	WTP reflects perceived benefit of screening	247	11.8
	Screening deemed worthwhile, given recognised benefits	20	
	Reassurance and peace of mind	71	
	Screening offers early detection of disease	140	
	Screening can save money for the NHS	16	
E8	Reported familial experience of colorectal cancer	65	3.1
E9	Protest expressed at the idea of payment	395	18.9
	Having paid taxes, one shouldn't have to pay more	161	
	Screening is vital and must be free	69	
	The NHS should bear the costs	124	
	The service should be free, to encourage use	27	
	Free tests will benefit the NHS by cost saving	14	
TOTAL		2087	

Additional supplementary explanations

E10	Relative acceptability of the tests, e.g. pain, inconvenience, intrusion	60
E11	Relative affordability	32
E12	Relative likely costs	87
E13	Relative benefits offered by the tests	63

Table 6.2 indicates that nine broad categories of explanation were finally identified. Four supplementary categories were created for respondents who had made additional and specific distinctions between FOB and FS. The data in this table comprise all the reasons offered from the entire sample, and therefore include comments from subjects that provided no or zero, as well as positive WTP values. Phrases that are in italics in table 6.2 represent examples of explanations made by the respondents. A total of 1,523 subjects offered a written explanation, and of this group, a single coding category seemed sufficient to encapsulate 979 (64.3%) explanations. However, for 423 (27.8%) subjects, two categories of explanation appeared to be present in the text, with three or more categories being necessary to describe the remaining reasons. Table 6.2 illustrates that affordability or ability to pay (E4) emerged as the most popular class of explanation, followed by expressions of protest either about the idea of monetary valuation or about potentially being obliged to pay (E9). The subjects that referred to acceptability of the WTP values are in a sense the opposite to protestors as they have indicated that the amount they are WTP is a fair or reasonable amount.

6.4.5 WTP for FOB and FS screening

The distribution of WTP values had skew coefficients of 22.3 and 39.0 for FOB and FS, respectively. The ranges of the values were considerable, the 25th percentile was £15 for FOB and £20 for FS and the 75th was £100, in each case. In fact, at the extreme, 9.3 and 11% of respondents offered zero WTP values for FOB and FS, respectively, whilst 4.3% and 2.7% respondents offered values equal to or more than £500.

i. Non responders and 'zero' WTP values.

Table 6.3 presents the results of two sets of logistic regressions that were used to firstly predict the characteristics of subjects not responding to the WTP question and secondly, identify the characteristics of subjects that revealed a zero WTP valuation. If respondents were not worried about colorectal cancer, if they found the WTP question inapplicable or difficult to answer, and if they did not offer an explanation for the FOB test, a 'no response' was more probable. The FOB and FS models are similar but have a few variations. In the FS model, the coefficient on WTP acceptability is no longer significant. The positive effect of higher incomes in the FOB model is matched by a significant negative effect of low income. If the respondent perceives themselves as having a higher than average chance of developing colorectal cancer then they are more likely to provide a value. Smokers are more likely to provide a response. Finally, as one would expect, an expressed preference for FS screening means that a response to the FS question is more likely, however a preference for FOB makes it less likely.

The results of the logistic regression designed to predict the non-responders appeal to intuition. It is likely that the significance of the age coefficient in the FS equation probably arises from the age condition in FS testing (respondents over the age of 60 feeling that the question is inapplicable to them). No response was significantly less likely if subjects had been offered the payment scale format. Earlier experiments with model formulation had included interactions between each of the reasons and format. None, however, achieved significance, supporting the view that the explanations offered operated independently of format.

Table 6.3: Logistic regression – predicting non-response and zero values

	Dependant: non-response (=1)						Dependent: zero WTP value (=1)					
	FOB			FS			FOB			FS		
	b	SE	Exp (b)	b	SE	Exp (b)	b	SE	Exp (b)	b	SE	Exp(b)
Format (payment scale =1)	-1.039	0.125	0.354	-0.597	0.110	0.551						
E1 (inapplicable)	2.145	0.308	8.538	2.642	0.307	14.034						
E2 (difficulties)	2.190	0.220	8.931	1.952	0.195	7.042						
E5 (acceptable value)	-1.565	0.528	0.209									
E9 (protest)	0.666	0.230	1.946	0.427	0.192	1.532	1.951	0.182	7.037	1.870	0.175	6.489
No reason given	1.920	0.170	6.821	1.623	0.136	5.067						
Gender (male = 1)							1.292	0.176	1.339	0.450	0.172	1.568
Age				0.015	0.004	1.015				-0.019	0.006	0.982
Income 0-£10,000				0.295	0.134	1.343	0.637	0.193	1.890	0.654	0.194	1.923
Income £20-30,000	-0.658	0.169	0.518	-0.422	0.157	0.656						
Income >£30,000	-0.461	0.181	0.631	-0.486	0.176	0.615						
Age leaving education							-0.049	0.026	0.952	-0.033	0.022	0.967
Visits to dentist	-0.141	0.035	0.868	-0.069	0.033	0.934	-0.105	0.052	0.900	-0.111	0.050	0.895
Health motivation score	-0.269	0.090	0.764	-0.210	0.085	0.810						
Current smoker (1=yes)				0.294	0.134	1.342						
Not worried about colorectal cancer	0.372	0.122	1.451	0.311	0.112	1.365	0.701	0.483	2.015	0.668	0.167	1.950
Chances above average				-0.355	0.188	0.701						
Preference for FOB	-0.829	0.126	0.436	0.223	0.114	1.250						
Preference for FS				-0.678	0.190	0.508						
Constant	-0.457	0.399		-1.708	0.464		-2.653			-1.942	0.521	
Pseudo-R-squared	0.310			0.270			0.152			0.153		

The FOB and FS models were also very similar when the logistic regression was conducted to predict the characteristics of subjects providing a zero valuation. If the respondents visited their dentists more frequently and possessed more years of education then a zero value was less likely. On the other hand, with the FOB model, protestors, males and subjects on low income along with those who had expressed no worries about colorectal cancer were more likely to give a zero WTP value. With the FS model, once again, age is a predictor, this time, it is a predictor of not providing a zero value. This is hypothesised to be the same reason as above, that is, if the responder is over a certain age then they are inclined to offer no response rather than a zero response to signal perceived irrelevance.

ii. Positive WTP values

To explain the positive distribution of WTP values, a linear regression model was used. Given the skews and ranges, transformation to natural logarithms seemed appropriate prior to analysis (Altman 1991). When conducting parametric analyses it is important to confirm that the data originates from a population that is normally distributed. Such transformations reduce the influence of outlying (and thus atypical) values on the results of the analyses. Table 6.4 presents the results of the linear regression models where WTP is the dependant variable and all other independent variables have been entered and either rejected or chosen according to the stepwise selection process.

Once again, the FOB and FS formulations are similar. The payment scale method yields higher WTP values on average than does the open-ended instrument. This is an interesting result as the payment scale method has been shown to produce higher

Table 6.4: Linear regression for positive WTP values (natural logarithms)

	FOB		FS		Marginal impact (%)	
	b	SE	b	SE	FOB	FS
Format (payment scale =1)	0.246	0.057	0.561	0.052	27.8	75.3
E2 (difficulties)	0.386	0.112	0.234	0.079	47.2	26.3
E7 (benefit)	0.440	0.088	0.330	0.078	55.3	39.0
E8 (family experience)	0.505	0.179	0.514	0.160	65.6	67.2
E9 (protest)	-0.309	0.082	-0.340	0.074	-26.6	-28.8
Gender (male = 1)	0.171	0.061			18.7	-
Income 0 - £10,000	-0.212	0.077	-0.157	0.071	-19.1	-14.5
Income £20 – 30,000	0.186	0.076	0.171	0.068	20.4	18.7
Income > £30,000	0.423	0.081	0.420	0.073	52.6	52.3
Age leaving education			0.020	0.005	-	-
Visits to dentist	0.049	0.018	0.056	0.016	-	-
Health motivation score	0.101	0.051			-	-
Not worried about colorectal cancer			-0.129	0.057	-	-12.1
Chances above average	0.370	0.087	0.234	0.079	44.8	26.3
(constant)	2.965	0.224	2.938	0.112	-	-
Adjusted R-squared	0.104		0.162			

WTP values compared to the open-ended technique in previous health care research (Donaldson, Thomas et al. 1997).

There is a positive association between expressed WTP and ability to pay, it is well known that income has a positive effect on WTP and the results from this study are consistent with these previous findings (Donaldson 1999). The income effects are exactly what theory would predict with respondents drawn from the lowest household income class stating significantly lower WTPs and the opposite being the case for the higher income classes. It is clear from the results that the WTP values are also positively influenced by expressed difficulties in reaching a WTP value, perceived benefits from screening, having familial experience of colorectal cancer, perceiving a higher-than-average chance of contracting the disease and visits to the dentist. In the FOB model, if the subject were male and had a high health motivation score then that would predict a higher WTP value. Alternatively in the FS case, having no worries about the disease lowers WTP, whilst having experienced more years of formal education increases it. Once again, these results appeal to intuition, although perhaps less so in the cases of gender and experiencing difficulties in reaching a WTP value. However it is interesting that male gender and regular visits to the dentist were positive predictors of actual compliance in the UK FS trial (Sutton, Wardle et al. 2000), whilst dental visits and higher socio-economic class were positive predictors of actual compliance in the Nottingham FOB trial (Neilson and Whynes 1995).

Along with the results of the linear regression analysis, table 6.4 also displays the marginal impacts of the individual, binary factors on predicted WTP, calculated by comparing the estimated WTP with in turn, each of the dummy variables at unity and zero and all other variables at their mean values. For example, in the FOB model, setting the format at 1 and 0 in place of the format mean produces predicted WTPs of £54.5 and £42.6, respectively. What can be inferred from this result is that the payment scale format inflates WTP by 27.8 per cent, compared with the open-ended format. Looking at the corresponding estimates for FS produces a more extreme result, at £61.6, £35.1 and 75.3 per cent, respectively. As is evident from Table 6.4, the specific impacts of family experience of colorectal cancer, high income and a high level of perceived risk on stated WTP are substantial.

iii. Overall WTP data

Table 6.5 summarises the WTP information for the FOB and the FS test for each elicitation format. The data is expressed in trimmed means (which excludes 5% of observations equally at the two extremes of the distribution). The trimmed means show that the true mean is influenced significantly by the small numbers of very high WTP values.

Table 6.5: The WTP for FOB and FS by elicitation format

				Percentiles		
	Mean	5% trimmed	mode	25	50	75
Open-ended method						
FOB	129.6	66.3	50	10	30	100
FS	86.1	43.4	50	10	30	50
Payment scale method						
FOB	93.1	68.3	50	20	50	100
FS	90.8	70.1	100	20	50	100

Within the sample, 1,718 respondents (62.1 per cent) offered a non-zero WTP for both screening options. Of these, 826 (48.1 percent) offered the same WTP value for both FOB and FS.

iv. Relative WTP

Since as many as 48.1% of respondents offered the same WTP value for both tests, a logistic regression model was used to establish the various factors that discriminate between respondents who reveal the same WTP values versus those who reveal different WTP values for either of the tests. Therefore since there was a comparison between the values provided for the tests it seemed logical to include the relative explanations offered (E10 – E13, table 6.2) for preferences for either tests. The results of the logistic regression are displayed in table 6.6.

Table 6.6: Explaining differences between WTPs of the two screening options

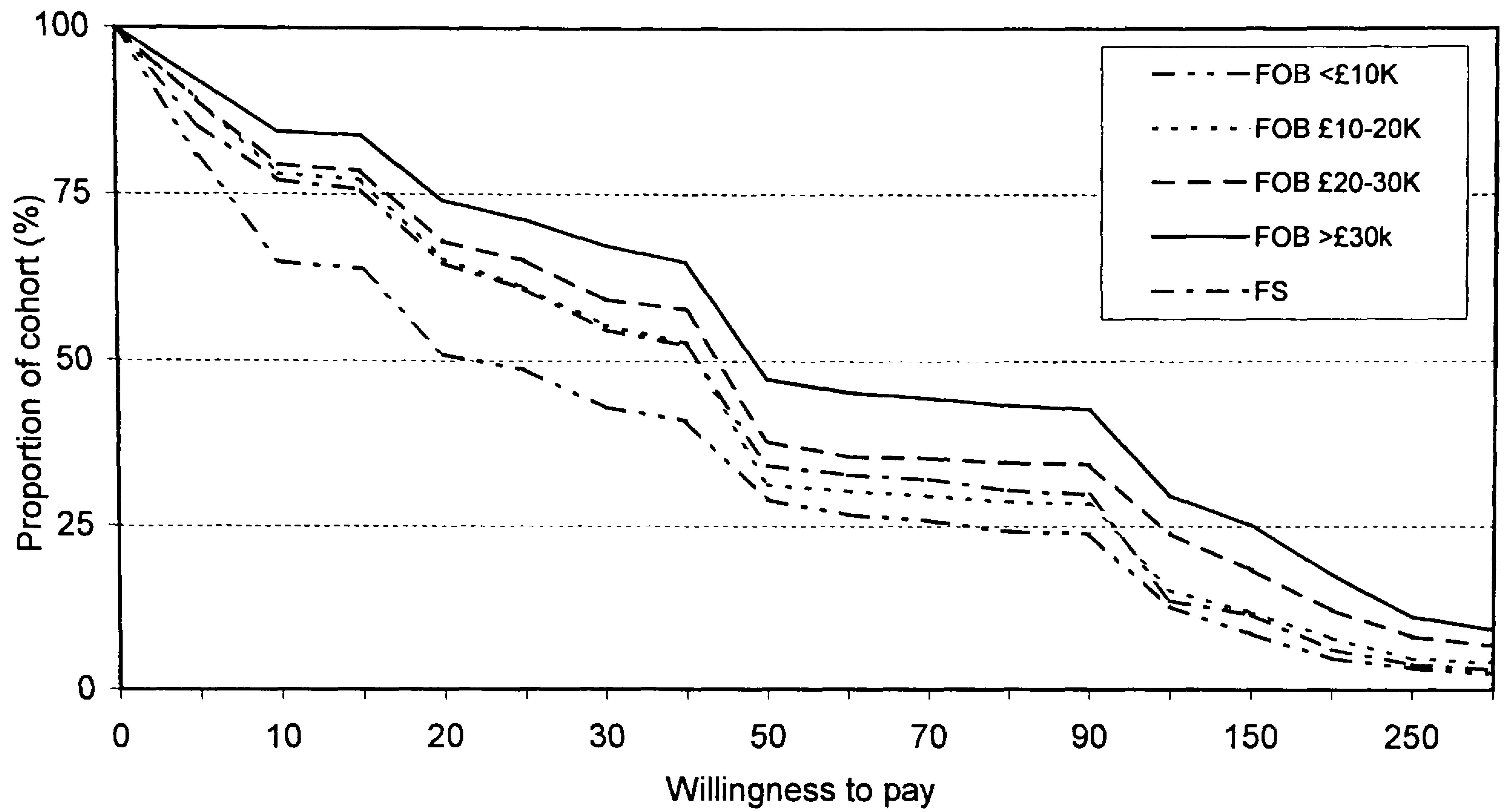
	Logistic regression		
	Dependent: if WTPs different (=1)		
	b	SE	exp (b)
Format (payment scale)	-0.582	0.110	0.558
E2 (difficulties)	0.697	0.227	2.007
E10 (relative acceptability)	1.687	0.552	5.405
E11 (relative affordability)	2.807	1.035	16.560
E12 (relative costs)	2.978	0.594	19.645
Income £20-30,000	0.278	0.134	1.320
Income > £30,000	0.565	0.150	1.760
Health motivation score	0.203	0.096	1.226
Constant	-0.767	0.406	
Pseudo / Adjusted-R-squared	0.143		

The payment scale format was more likely to produce the same WTP values for FOB and FS. However if the subject had expressed difficulties in reaching his or her WTP values, was from two of the higher income brackets, highly motivated and offered relative acceptability, affordability and cost as an explanation then different WTP values for the two tests were more likely. The inclusion of interaction terms in the model, between format and type of explanation, failed to yield additional significant coefficients.

6.4.6 Demand curves for FOB and FS.

Using the WTP data, it is possible to plot out demand curves for the FOB and the FS test separately. Figure 6.2 displays the demand curves for the two screening methods. Both the open-ended and the payment scale data have been combined to generate the demand curves. The graph shows the demand curves for the FOB test disaggregated by household income, to display the income effect, and for the FS test for the full sample. Each demand curve has the appearance of a demand curve that is frequently encountered for normal goods providing reassurance that the WTP demand curve for colorectal cancer screening is not particularly unusual. For the FOB test, the demand curves by income group are again as theory would predict, with the higher income group displaying a demand curve that lies above and to the right of the demand curve for the lower income group. It is evident that the demand curves and the median WTP values for both tests are very similar.

Figure 6.2: Demand for colorectal cancer screening, FOB (by income group) and FS (full sample).



PART B

6.5 Zero values

As already mentioned, the distribution of WTP values had skew coefficients of 22.3 and 39.0 for FOB and FS, respectively (a normal distribution has zero skew). Zero bids can be interpreted in four different ways (Kidholm and Langkilde 1994). The most common is known as a protest zero. Diamond and Hausman (1993) interpret it as a:

'wtp answer of zero dollars, given because a respondent wishes to make a protest against the payment vehicle or some other aspect of the survey, not because the respondent truly places a zero value on the good being valued'.

Alternatively, a zero bid can be interpreted as a negative zero. In this case the respondent may benefit from an increase in risk (not having the screening test) however the wording of the question does not allow for a negative bid. Negative zeros should not be included in the analysis, as inclusion will result in a positive bias when calculating the average WTP.

A zero bid will be interpreted as a positive bid when the respondents explicitly state that they find the actual decision process costs too much. In this case allowance will be made for the fact that this group of respondents may have a potentially positive WTP when calculating the average WTP.

And last, a zero bid can be interpreted as a true zero bid. Here the respondent thinks that the marginal benefit of a further risk reduction (having the screening test) is zero. Since this class of zeros are termed the 'true zeros' it is advised that they be left in for the overall analysis.

In this study, the reasons provided by the respondents for their WTP for either FOB or the FS test were used to interpret the zero valuations. Table 6.7 displays the results:

Table 6.7: Zero bids for FOB and FS

	Open-ended		Payment scale	
	FOB	FS	FOB	FS
Zero values	83	92	117	126
Zero values with reason:	44	53	79	86
Positive zeros	4	6	0	2
True zeros	1	1	1	3
Protest zeros	38	43	78	81
Negative zeros	1	3	0	0

A total of 418 zero bids were elicited across both question formats for both of the tests. Of these 418 values, 262 zero bids were accompanied with a reason enabling an interpretation of the zero value. Results of the analysis show that 92% of the zero bids are classed as protest zeros, 5% as positive zeros, 2% as true zeros and finally 1% as negative zeros. According to theory (Kidholm and Langkilde 1994), all true and positive zeros should be kept in for analysis and negative and protest zeros should be omitted. The effect of this exclusion criterion will be dealt with later in the chapter.

6.6 Inconsistent responses

Logic predicts that to be consistent, all respondents must provide a greater WTP value for their most preferred test over their least preferred test. By looking at the direction of WTP values, for each of the preference groups, the consistency of the WTP results can be measured. Only respondents who elicited a WTP value for both tests and provided a response to the 'preference question' (i.e. prefer FOB, FS or no preference) are included in the analysis of consistency. Therefore a total of 1902 respondents are to be analysed. Table 6.8 displays the results:

Table 6.8: WTP for preference groups

	Is WTP FOB > WTP FS?			
Preference Group	No (%)	Equal (%)	Yes (%)	Total (%)
FOB	199 (10)	433 (23)	200 (10)	832 (44)
FS	82 (4)	125 (7)	49 (3)	256 (13)
No Preference	184 (10)	442 (23)	188 (10)	814 (43)
TOTAL	465 (24)	1000 (53)	437 (23)	1902 (100)

The table shows that 43% of the respondents had no preference between FOB and FS therefore this group of respondents are labeled as 'no preference'. What is interesting is that ten percent of individuals that preferred the FOB test were WTP more for the FS test and three percent that preferred the FS test were WTP more for the FOB test. These results seem inconsistent with what one would expect. Donaldson (1997) suggests that 'inconsistent responses' arise through subjects attempting to estimate the resource cost of the programme that they are being asked to evaluate. Given that the open-ended question design is thought to produce more subjects estimating cost, this problem is accentuated when subjects are answering the open-ended question format (Donaldson, Thomas et al. 1997). Miguel and Ryan (1998) however found that the problem of an inconsistent response is just as likely with the payment scale format

(Ryan and Miguel 2000). To try and understand why an inconsistent response may arise, the qualitative comments provided by the respondents are examined. These explanations were grouped into thirteen categories described in table 6.2 and one reason provided by one subject may be classed in more than one category. Table 6.9 presents the reasons provided by the respondents.

Table 6.9: Reasons for 'inconsistent' WTP values

Type of explanation provided		Question design	
		Open-ended	Payment Scale
E1	Question deemed inapplicable	2	0
E2	Evidence of difficulty in estimating WTP	6	16
E3	WTP based on nominal amount	3	2
E4	WTP reflects ability to pay (affordability)	29	46
E5	WTP reflects a fair, acceptable or reasonable value	10	17
E6	WTP reflects costs of screening	15	17
E7	WTP reflects perceived benefit of screening	18	16
E8	Reported familial experience of colorectal cancer	9	1
E9	Protest expressed at the idea of payment	115	133
E10	Relative acceptability of the tests	4	3
E11	Relative affordability	3	2
E12	Relative likely costs	16	13
E13	Relative benefits offered by the tests	5	8

As table 6.9 reveals, an approximately equal number of respondents referred to the resource cost of the tests for the open-ended and the payment scale designs. Contrary to Donaldson's findings therefore there was no real propensity to mention resource cost when answering the open-ended format. In fact, the results of this study show that more subjects referred to resource cost with the payment scale design. In addition to more subjects mentioning resource cost, a greater number of subjects also referred to their ability to pay, made protest comments and indicated that they found the WTP question difficult to answer when completing the payment scale format (compared to the open-ended format).

Plausible reasons behind an inconsistent response suggest that an attempt to estimate the resource cost of the test and difficulties in understanding the WTP task can increase the likelihood of inconsistency occurring. Based on the definition of an inconsistent response, 248 subjects in this study provided inconsistent answers of which a total of 509 'categorised' explanations for these answers were provided. Only four percent of these comments referred to difficulty in answering the WTP question and 6% mentioned resource cost. The majority of 'inconsistent' comments in this study were protest comments (49%) and a substantial proportion referred to affordability issues (15%). According to theory, all inconsistent respondents should be removed from further analysis; the effect of this will be examined later in the chapter.

An important point to mention however is that it is plausible that these inconsistent responses arose through random error whereby the respondents answered in an inconsistent manner due to lack of concentration, attentiveness or 'by mistake'. Harless and Camerer incorporate an error rate into their analyses of different theories to explain violations of expected utility theory in order to take account of systematic variation in responses that are inconsistent (Harless and Camerer, 1994).

Assume that in both choice and valuation tasks that there is a probability e that individuals will behave contrary to their true preferences. This means that in the choice task for example, and assuming that true preference is for the FOB test, that the probability that the FOB test will be chosen is 1-e and the probability that the FS test will be chosen is e. Taking both choice and valuation tasks together, there is therefore a 2e(1-e) probability of responses being inconsistent between the tasks in either direction.

We know from the above results that 10% of the sample chose FOB as their preferred test yet gave a higher value for the FS test, and that 3% of the sample chose FS as their preferred test yet gave a higher value for the FOB test. It therefore follows that if 2e(1-e) = 13% then the error probability within this study is equal to 7%. Thus using the Harless and Camerer account of behaviour, there is a 7% probability that individuals in this study will behave contrary to their true preferences.

6.7 Intensity of preference

The relative difference between the WTP values helps to provide information on the intensity of preferences for either test. To measure the intensity of preference, the WTP ratios for each of the preference groups are calculated. Therefore equal WTP values for each test will suggest that there is no preference for either test. Equal WTP values were elicited from 23% of respondents that preferred FOB and 7% that preferred FS. It is plausible that respondents might not have a strong enough preference for their preferred test to provide a higher WTP for it but why then do they not indicate 'no preference' in the preference direction question. If they had a slight preference for one test then theory suggests a slightly higher WTP value should have been elicited for the FOB test over the FS test. The calculated WTP ratios for each of the preference groups are illustrated in table 6.10. Here the analysis concentrates only on the group of subjects that answered both the preference direction question and gave a WTP value for both tests (n=1902).

Table 6.10: WTP ratios for the FOB and FS test.

Preference group	WTP ratio
All cases	WTP FOB/ WTP FS = 1.31
FOB test	WTP FOB/ WTP FS = 1.34
FS test	WTP FOB/ WTP FS = 1.18
No preference	WTP FOB/ WTP FS = 1.33

Table 6.10 has produced a surprising result. Overall, the mean WTP for the FOB test is 31% greater than for the FS test. This result is what one would expect as it corresponds with the proportion of respondents in each of the preference groups (the majority prefer the FOB test). However if the sample is divided into the appropriate preference groups and the ratio of mean WTP elicited for each test examined, discrepancies arise in the data. As expected, subjects who prefer the FOB test are WTP 34% more for their preferred test compared to the FS test. However surprisingly, subjects who prefer the FS test are willing to pay 18% more for the FOB test; reasons why this result has occurred are unclear. The no preference group are also WTP more for the FOB test by 33%. Again, attention has to be drawn to a possible inconsistency in the data. It is illogical for an individual that prefers the FS test to reveal a higher WTP for the FOB test and vice versa. Possibly the respondents have simply misunderstood the question or alternatively, resource use considerations have been encountered, e.g. despite preferring the FS test, the respondents may feel that perhaps they are getting 'more value for money' from the series of FOB tests. The only means of getting closer to solving these inconsistencies is to examine the

qualitative comments provided by these individuals. As mentioned before (table 6.8), the 49 respondents that preferred the FS test produced a greater mean WTP for the

FOB test. Sixty-nine percent of this group provided explanations for their WTP response. Only two respondents had difficulty dealing with the WTP task. However, 26% referred to their ability to pay (affordability) and 24% attempted to estimate the resource cost of screening. The most striking finding, from examining the qualitative data obtained from this group, is that all respondents that provided an explanation registered a protest vote. It is clear from this result therefore that the reason for the discrepancy in the direction of WTP values is that nearly three-quarters of respondents in the group are classed as protestors. Again, the protestors should be omitted from the analysis and the effect of this will be seen in the exclusion criteria section below.

6.8 Exclusion criteria

As discussed, if a subject elicits a protest vote or displays inconsistencies in their response then it is recommended that they be omitted from the overall analysis of the WTP data. Table 6.11 displays what happens to the 25th, 50th and 75th percentiles when this happens.

Table 6.11 – Exclusions in the WTP results (£, quartiles)

	Open-ended				Payment Scale			
	25%	50%	75%	n=	25%	50%	75%	n=
FOB Test								
NONE	10	30	100	955	20	50	100	1201
Difficulties / guess	10	30	100	838	20	50	100	1077
Protest	10	40	100	838	20	50	100	972
Inconsistencies	10	50	100	329	20	50	100	395
ALL	15	50	100	247	20	50	100	294
FS Test								
NONE	10	30	50	895	20	50	100	1084
Difficulties / guess	10	25	50	792	20	50	100	984
Protest	15	30	59	779	20	50	100	870
Inconsistencies	10	25	50	329	10	50	100	395
ALL	15	30	50	247	20	50	100	294

Table 6.11 presents the WTP data for both of the tests and the question designs. The first row in each box represents the estimates for the full sample without any exclusions. Each subsequent row displays the estimates as more data are excluded based on the set criteria discussed previously. The table clearly shows that removing the data from analysis makes no or very little difference to the overall median WTP (50^{th} percentile). In the case of the payment scale format (FOB test) the exclusions have absolutely no impact. The outcome in table 6.11 casts doubt as to whether excluding respondents who are classed as inconsistent or protestors is really required if your sample size is large enough.

6.9 Discussion

The results have shown that the range of WTP values elicited for the FOB and FS tests were considerable. The distributions were highly skewed with the median and the mode of each distribution being £50 and the second and the ninth percentile being

£10 and £200 respectively. A few extreme values were elicited as 4.3 per cent and 2.7 per cent of the sample offered WTP values equal to or more than £500, for the FOB and the FS test respectively. The skewed distribution is something that is common to most WTP studies - within a given sample, it is common to find a small number of very large WTP values that influence the average values.

It is important to understand the reasoning behind the WTP responses therefore the qualitative aspect of the study is essential. Asking the subjects to describe the reason why they chose a particular WTP value for the test provides insight into the thought process that they went through whilst completing the WTP task. This helps to distinguish between the individuals that have a genuine demand for the test and the individuals who have misunderstood the WTP task that has been set out for them. Affordability emerged as being the single largest class of explanation for the WTP values in this study; this was closely followed by protest comments. Referral to ability to pay seems a logical response when considering willingness to pay. After all, the subjects are encouraged to consider the opportunity cost of their maximum WTP value. As long as the ability to pay is equally distributed across the study sample (representative sample in each income group) then consideration of 'affordability' issues does not render a problem for analysis. How to deal with the protestors is not quite as straightforward. Protestors are subjects who have objected to the style of questioning and have responded accordingly. It is difficult to ascertain whether the WTP values revealed from this sample are 'true' WTP values. According to theory, protestors should be removed from further analysis but the results of this study have shown that removal of this group makes no or very little difference to the overall distribution of WTP values. In fact with the payment scale question design, for the FOB test, it made no difference whatsoever.

The WTP values revealed for both tests are very similar. The median WTP for FOB and FS amounts to £50 for the payment scale type of questioning and £30 and £50, respectively for the open-ended style of questioning. It is therefore clear that the only factor that is having an impact on the WTP values is not the different aspects of the test but the style of questioning. Here, it is possible that what may have been encountered is the 'embedding effect' (Kahneman and Knetsch 1992), a feature that has been recognised in the WTP literature. Instead of considering the different processes of care and aspects of each test the respondent has provided a value for colorectal cancer screening in general. Alternatively a 'framing' effect may have been encountered (Mitchell and Carson 1989). When answering the open-ended question design the respondents are asked to 'think of a number' with no further guidance. In contrast, the payment scale question asks the subjects to 'choose a number' in a defined range, using a specified decision-making process. The payment scale question therefore gives the subjects a frame of reference to work from which is conceptually easier to respond to and easier to complete than the open-ended design. Also the subject knows, with the payment scale design, that their response can legitimately lie within the pre-determined range offered. The chosen range will include values that is well above the range of WTP amounts expected from the sample therefore causing the average WTP (from the payment scale design) to be pushed upwards. Obviously this means that the payment scale result will always be susceptible to the range of values (range bias) chosen for the scale (this is an issue dealt with in chapter seven). The different results from the open ended and payment scale question formats has important implications for the design of future WTP studies as the choice of question design obviously has a strong influence on the final WTP values that are revealed.

The study results show that 'no response' is more likely with the open-ended format, that a difference between the FOB and FS WTP values are less likely with the payment scale and that the payment scale yields a higher WTP value on average. These results are consistent with previous findings that have shown that the payment scale design produces a greater response rate and a higher WTP on average (Donaldson, Thomas et al. 1997). Alternatively, the study results have shown that the payment scale design does not produce fewer zero valuations and protest valuations, a claim made by previous findings (Klose 1999). And in comparison with other studies (Donaldson, Shakley et al. 1997; Donaldson, Thomas et al. 1997; Ryan and Miguel 2000), relatively few of the subjects offered estimates of the resource cost of screening as an explanation for their WTP values. Therefore the study results do not support the hypotheses advanced in these studies, that subjects offering WTP values based on their assessment of costs necessarily accounts for either the different results produced by open-ended and payment scale formats or observed inconsistencies between WTP values and preferences.

'Inconsistency', with respect to a violation of expected utility theory, is not something that is uncommon and confined to contingent valuation as it has been observed repeatedly in other areas of health economics such as the valuation of health state utilities (Dolan and Kind 1996; Krabbe, Essink-Bot et al. 1997; Badia, Roset et al. 1999). The assessment of consistency is based solely on internal logic, i.e. analysis within the dataset taken at face value. Sen (1993) however argues, that 'there is no way of determining whether a choice function is consistent or not without referring to something external to choice behaviour (such as objectives, values, or norms)' (Sen 1993). What Sen is saying is that observed inconsistency within an experiment might be illusory and might not exist if only a better experiment had been conducted, to obtain more information from the subjects. Alternatively, it is quite conceivable that observed inconsistencies between preference and value might be less the aberration which orthodox economic theorists consider it to be, and more an essentially normal human characteristic. Preference-reversal – whereby individuals declare they prefer A to B yet offer to pay more for B than for A – has been identified by experimental economists since the 1960s and appears to be a widely prevalent phenomenon (Roth 1995). This was discussed in chapter four. Explanations for reversal vary, from cognitive illusion, analogous to optical illusion, to the belief that establishing preferences and establishing values actually entail the use of different mental processes which only partially map onto one another (Heap, Hollis et al. 1992). In consequence, many economists now see the orthodox experiment for preference-value consistency as a substantial obstacle to the analysis of real-world behaviour (Starmer 2000). By implication, and given the observed prevalence of response inconsistency, the rejection of inconsistent responses in value elicitation would entail an analysis of a sub sample of individuals that may be quite atypical of the population from which they were drawn.

The study set out to establish the demand for FOB and FS testing for bowel cancer using the WTP methodology. The results show that individuals who are from a low income group, do not visit their dentist regularly and have a low health motivation score are not likely to demand colorectal cancer screening. Individuals from a high income background, who perceive screening as being beneficial, have familial experience of colorectal cancer, perceive a higher than average chance of contracting the disease and visit their dentist regularly have a positive WTP for colorectal cancer screening. All these results appeal to intuition.

The purpose of a WTP study is to determine the value that is placed on a good or programme that is being investigated, in this case, colorectal cancer screening programmes. Translating a value from a hypothetical setting to an actual setting is always going to be problematic especially if the method to measure this value (like WTP) is far from established. However in the case of colorectal cancer screening some additional information exists that helps to interpret the results. First, the Flexi-scope trial for FS screening undertook a sub-study set out to measure the time and travel costs associated with a clinic visit at hospital to receive a FS examination. Approximately 3500 subjects were sampled from 12 sites in the UK and median time and travel costs were estimated at £19 per attendance, within an inter-quartile range of £11.8 to £29.4 (Frew, Wolstenholme et al. 1999). These values can be interpreted as a proxy for revealed preference figures (not a true revealed preference figure as this method does not take account of respondents who decline to attend) for FS screening and are of the same order of magnitude as the average WTP values elicited in this sample, although somewhat lower. Second, neither FOB or FS screening has been fully implemented into the NHS however estimates of the resource cost entailed by both protocols have been made. The 1989 cost of three rounds of screening, using the Nottingham trial protocol, was estimated to be £14.3 (Walker, Whynes et al. 1991). Given the age range for eligibility, the maximum number of rounds a person might be screened under the protocol as presented in the WTP study would be thirteen, implying a total programme cost of around £62 per subject. The cost of a sigmoidoscopy investigation was estimated at approximately £56 in 2002 (Whynes, Frew et al. 2002). Even after allowing for inflation over the past decade, it is evident that the median WTP values expressed by our sample would be quite similar to the probable cost of the services, were they to be introduced.

A major objective of this study, other than an investigation of the demand for colorectal cancer screening, was to ascertain how the question design affects the WTP values. As a guide to the reader, table 6.12 summarises the main findings from this first piece of empirical research. As discussed, the results are consistent with the findings of some similar studies but contrary to others. Chapter four described all the different elicitation formats that can be adopted when conducting a WTP test, the choice of format does not just centre on open-ended versus payment scale. The results of this study encouraged the later research around the performance of other elicitation designs that are discussed in the forthcoming chapters. One aspect of the payment scale design that seems fundamental to the magnitude of WTP values elicited concerns the range of values chosen in the scale offered to respondents. This was an issue that was felt to be necessary to address and the result of the investigation is the subject of the next chapter.

Table 6.12: Summary of findings from chapter six.

Findings
1. Distribution of WTP values were highly skewed for both question formats.
2. The FOB test was the overall preferred test with 46% of the sample choosing it over FS (14%) and having no preference between both tests (39%).
3. Overall, affordability or ability to pay emerged as the most popular class of explanation for the WTP values.
4. The payment scale question yielded a higher response rate and higher WTP on average compared to the open-ended question format (£30 (op) vs £50 (ps)).
5. There was a positive association between expressed WTP and ATP.
6. 48% of the respondents offered the same WTP value for both FOB and FS.
7. The demand curves had shapes that are frequently encountered for normal goods suggesting that the WTP data collected for colorectal cancer screening is not particularly unusual.
8. 92% of the zero bids offered (with explanation) were classed as protest zeros.
9. 248 subjects in the sample provided inconsistent responses.
10. The majority of explanations offered for the inconsistent responses relates to the affordability and protest comments.
11. Overall, the mean WTP for the FOB test is 31% greater than that for the FS test.
12. However, subjects who preferred the FS test provided an average WTP that was 18% more for the FOB test suggesting an inconsistency in preferences.
13. 24% (of those offering an explanation) of this FS 'inconsistent' group attempted to estimate the resource costs, 26% referred to their ATP and all respondents recorded a protest value.
14. After exclusions were made based on inconsistencies, protestors and those who had difficulty with the task, the WTP for FOB and FS altered slightly for the open-ended question and did not change for the payment scale question.
15. The WTP for FOB and FS are very similar, the biggest influence on the WTP is the style of elicitation format, not the different aspects of each test.

Chapter Seven - Range Bias

7.1 Introduction

Despite the advantages that the contingent valuation (CV) method holds, the application of the willingness to pay (WTP) method to measure benefits in health care studies is far from refined. Chapter four discussed many issues relating to the methodological quality of the technique, one of which concerns the choice of elicitation format. There is considerable debate over the precise style of question that should be applied when eliciting WTP values essentially because each of the formats appear theoretically-prone to different types of bias.

Chapter six in the thesis describes a study designed to elicit WTP values for colorectal cancer screening using the open-ended and the payment scale technique. The focus is on looking at the relative valuations revealed for faecal occult blood (FOB) testing versus flexible sigmoidoscopy (FS) testing, two alternative screening programmes for colorectal cancer. Briefly, these results showed that the choice of question format produced a different set of results, the payment scale format achieved a greater response rate and a higher average WTP value.

The payment scale approach, developed by Mitchell and Carson (Mitchell and Carson 1981; Mitchell and Carson 1984), invites subjects to select their maximum WTP from a specified and ordered list of possible values. Mitchell and Carson developed the design as an alternative to the iterative bidding game (Davis 1964) and the open-ended approach. Both the iterative bidding design and the open-ended method have been discussed as having their own set of biases. The iterative bidding game is thought to be susceptible to 'starting point' bias (this effect is investigated in chapter nine). Whereas the open-ended approach, because respondents are unguided and unprompted, can leave the respondent feeling unfamiliar with the process. Consequently this may lead them to refer to resource costs of the test being valued, or fail to comply with the task completely.

Mitchell and Carson designed the payment-scale question with the aim of maintaining the direct question approach with the avoidance of starting from a specific point. The design means that respondents are asked to choose their maximum WTP value from a given range of potential values that are usually presented from the lowest to the highest amount. Mitchell and Carson felt that the respondents therefore are presented with a context of values to choose from but are not asked a specific value (Mitchell and Carson 1989). The payment scale format has been widely used in health economic evaluations (Diener, O'Brien et al. 1998; Smith 2000) both for the reasons which Mitchell and Carson have advocated and for the reason of cost and convenience to researchers. It is easy to adopt such a question format when using postal questionnaires as the data capture instrument. However the payment scale method is prone to range and central point bias. Subjects confronted with a scale to choose from can regard that scale as representing 'reasonable' upper and lower bounds for valuation. In other words, respondents with an original low WTP valuation might take a long scale (extending to high values) as indicative that they 'ought' to value the intervention more highly. At the opposite extreme, those initially considering very high WTPs (beyond the scale length) might presume that such values would be considered 'unreasonable' and would thus revise their estimates downwards. Mitchell and Carson do admit that subjects may be influenced by the ranges used in the cards and the locations and benchmarks of the values in the range (Mitchell and Carson 1989).

These form of biases have important implications when comparing results elicited using the payment scale technique to those revealed from other question formats. The open-ended/payment scale study, discussed in chapter six is one such study. If bias does indeed exist and other things remain equal, stated WTP values will be higher if subjects are offered a longer payment scale, i.e. one with a higher maximum value. The potential existence of range bias has been discussed extensively in the literature but has been rarely tested in health economic studies (Klose 1999). The study described in this chapter aims to address this issue by confronting the question of range and central point bias directly.

7.2 Study Design

7.2.1 Background

As before (in chapter six), the study uses the WTP method to measure the value that people place on colorectal cancer screening. The focus is on looking at the relative valuations revealed for faecal occult blood (FOB) testing versus flexible sigmoidoscopy (FS) testing.

7.2.2 Data capture instrument

The data capture instrument in this study had a similar design to the one used in the original open-ended/payment scale study. To recap, the first two pages of the instrument contained a description of the study, colorectal cancer and screening in general. Subjects were then provided with detailed descriptions of each screening programme (FOB testing and FS examination). Once the subject had read through the descriptions they were then asked if they would like to have a screening test for colorectal cancer and if so, which screening programme would they prefer. Subjects were invited to respond by indicating FOB, FS or no preference. Once the direction of preference for the screening programme had been revealed the subjects were then asked for their maximum WTP first for the FOB test and then for the FS test. Using the payment scale style of question, WTP values were then elicited. Previously, in the original open-ended/payment scale study, the vertically arranged list of values ranged from top to bottom, £0, £5 and £10, thereafter to £100 in units of £10, to £200 in units of £20, to £500 in units of £50 and to £1000 in units of £100. Subjects expressing a valuation in excess of £1000 were requested to write in the appropriate amount (i.e. beyond the limit of the scale, the payment scale defaulted to the open-ended format). In this study these values were replaced by a different set of values, the original long scale from £0 to £1000 was replaced by a much shorter scale of £0 to £100, with increments of £5. The respondents were again given the option of writing down their maximum value should it exceed £100. The short payment scale questionnaire is presented in appendix five. With the insight of the results from the first open-ended/payment scale study, it was evident that the maximum value of £100 in the new shorter scale was twice the median WTP value (£50) that emerged from the original open-ended/payment scale study (where the long scale was adopted). Given that the maximum value from the first study was £1000 (hence 20-times the resulting median WTP value), the two scale lengths differed by an order of magnitude.

Once the respondent had completed the WTP question they were then asked for their reasons why they had chosen that particular value. The questionnaire then became identical to the original instrument. To recap, information was requested on a range of socio-demographic and economic variables such as age, gender, age on leaving formal education, employment status, household income, health motivation, GP and dental visits and attitudes towards screening.

7.2.3 Distribution of questionnaires

Similar to the original open-ended/payment scale study, the questionnaires were distributed through a general practice surgery, this time, located in Nottingham, England. The practice received a visit from myself at intermittent time intervals to clear up any problems concerning the study and to collect the completed questionnaires. As and when required, new uncompleted questionnaires were taken to the practice for the GPs to distribute – this was instead of giving the practice the questionnaires in bulk to distribute at the beginning (as in the open-ended/payment scale study). It was easier to follow this procedure as the general practice surgery was in close proximity to the research office making it convenient to visit regularly.

The questionnaires were handed out using the same exclusion criteria as before. The general practitioners (GPs) were asked to distribute them during normal consultations using an exclusion criteria, that is, to no one under twenty-five years of age, on the grounds of likely perceived irrelevance of screening in that age group. To minimise distress, subjects with a recent diagnosis of colorectal cancer in the family were to be excluded. And finally the GPs were asked to exclude anyone with substantial reading, learning or language difficulties on the grounds of incapacity to complete the questionnaire. The respondents were encouraged to complete the questionnaire whilst inside the general practice building and then to 'hand in' the completed questionnaire to the receptionist. It was thought that this method would increase the response rate. However, anyone with a strong desire to take the questionnaire home was provided with a pre-paid envelope to return the questionnaire to the university once completed.

The overall aim of this study was to compare the results obtained from the original payment scale (long-scale) project to that obtained from the payment scale question when a shorter scale is adopted. The detailed results of the long scale study have been reported in chapter six, however, for comparison with the shorter scale, the long scale results will be summarised and presented again in this chapter.

7.2.4 Analysis

The long-scale and short-scale instruments will be compared looking at the differences in response-rate to the WTP question. To check initially that there are no underlying differences in preference direction between the samples, comparisons will be made in terms of the preference groups. The qualitative data obtained from each of the scales will be examined. The long-scale and short-scale WTP values will then be analysed with respect to determining the impact of using different scales on the results. The primary hypothesis being that, in comparison with the original payment scale results, the shorter scale will compress valuations. To control for the effects of confounding variables that might act upon WTP independently, a regression analysis will be conducted by combining both the short and long scale data and using the WTP as the dependant variable. A dummy variable will then be used to distinguish between the short and the long scale instruments. Finally, the data will be used to map demand curves for FOB and FS screening by scale type.

7.3 Results

7.3.1 Response rate

In total, 202 completed questionnaires were received for the short-scale version to compare against the 1,401 replies for the long-scale version. Due to the difference in distribution protocol the overall response rate to the long-scale and short-scale study cannot be compared. The GPs received questionnaires as and when required with the shorter scale study compared to receiving them in bulk with the long-scale study. However the difference in proportion of individuals providing a response to the WTP question can be compared. A similar proportion of individuals failed to respond to both the short scale and the long scale version (12.6% and 11.9% for the long scale and short scale respectively). Of the respondents that answered at least one WTP question (either the FOB or the FS question), there were a similar proportion for the short scale and the long scale (11.6% and 7.9% for the long scale and short scale respectively), this difference in proportion is insignificant (χ^2 =2.72, p=.256).

7.3.2 Socio-economic characteristics

Characteristics of the study sample that answered at least one WTP question for the short scale and the long scale are presented in table 7.1.

Table 7.1 Characteristics of the short-scale and long-scale sample

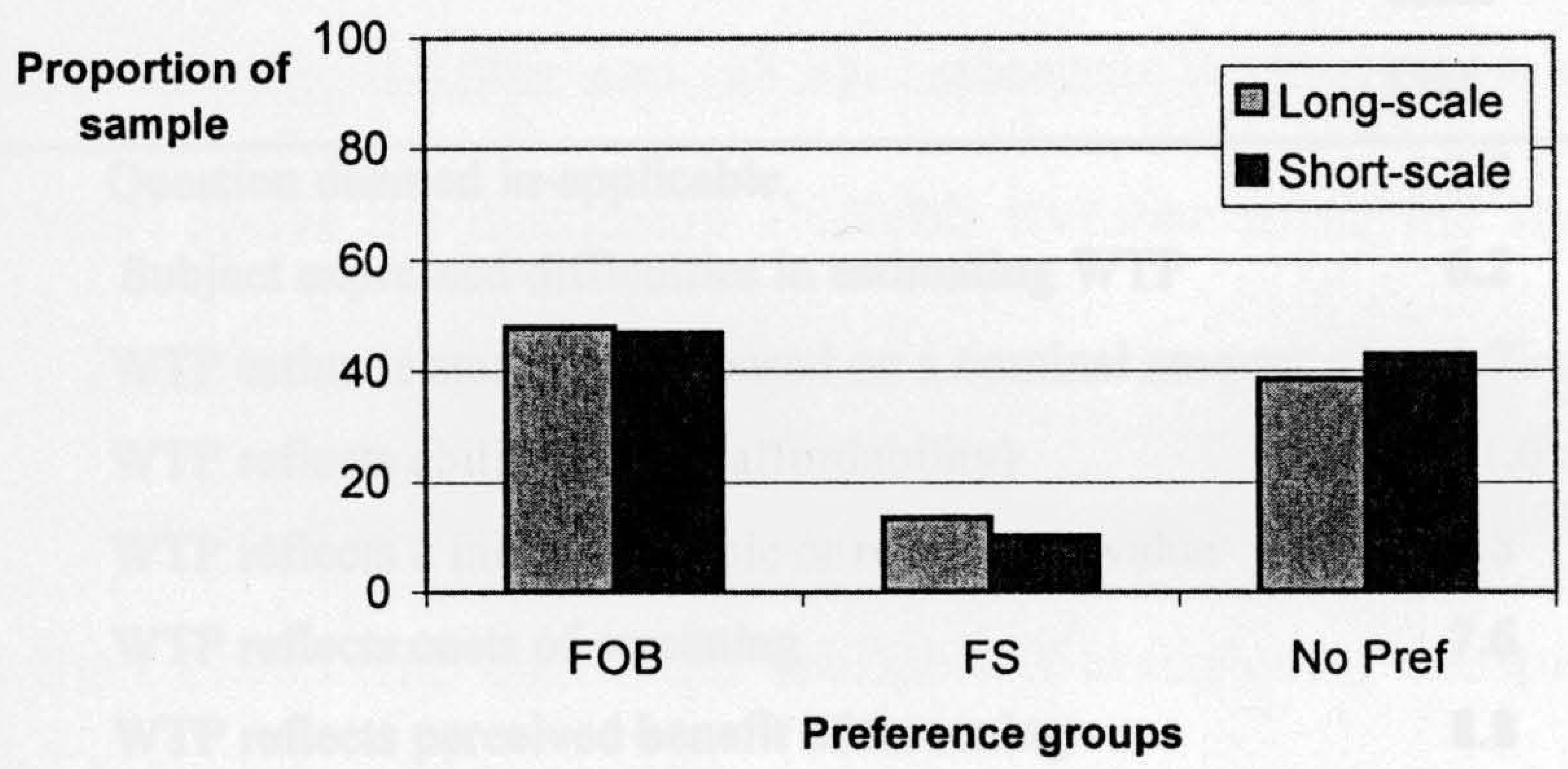
Variables	Long-scale	Short-scale	F or χ^2	p =
n =	1224	178		
Gender:				
Female	765 (63.3)	104 (60.1)	.647	.421
Age [median]	49 [37-60]	50 [39-59]	.059	.809
Proportion under 45 years	462 (40.5)	68 (41)	.012	.915
Income:				
< £10K	287 (26.4)	40 (27.6)	5.56	.135
£10-20K	362 (33.2)	39 (26.9)		
£21-30K	245 (22.5)	44 (30.3)		
>£30K	195 (17.9)	22 (15.2)		
Proportion quite/very worried about bowel cancer	226 (18.7)	47 (26.5)	9.00	.029
% who think chances of developing disease are higher than average	118 (10)	16 (9.1)	3.20	.668
Number of visits to dentist in last 2 years	4 [3-5]	4 [2-5]	2.11	.146
Age on leaving full time education [median]	16	16	.80	.370
Smokers (%)	228 (18.8)	46 (26.9)	6.19	.045
Health motivation score [median]	4.25 [3.75-4.5]	4.25 [3.5-4.5]	.34	.557

The proportions of individuals within each socio-economic characteristic in the long scale and the short scale are very similar. There are only two characteristics where the difference is statistically significant between the two scales these are firstly, smoking - there are a slightly higher proportion of smokers in the short scale group (27% v's 19%), and secondly, there are a slightly higher proportion of individuals who are worried about the disease in the shorter scale group (26% v's 19%). Although these differences are significant at the 5% level they are not at the 1% level.

7.3.3 Preference groups

Figure 7.1 illustrates the direction of preference groups for each of the scales.

Figure 7.1 Preference groups for long and short scale.



The above graph clearly shows that there are similar proportions of individuals within each scale that prefer FOB, FS and have no preference. The difference in the preference direction for each scale is not significant (χ^2= 2.11, p= .347). The majority of the sample for both scales chose the series of FOB tests as their preferred screening programme.

7.3.4 Qualitative information

The respondents were asked to explain how they chose a maximum WTP for each of the tests. The same coding frame (described in chapter six) developed to code the reasons obtained from the long scale sample was applied to code the reasons from the shorter scale sample. This coding frame evolved from each member of the research team reading several hundred responses selected at random, forming themes or categories and then reworking these themes with other team members to rule out any inconsistencies. Table 7.2 details the explanations obtained from the sample, for comparison, both the long scale and short scale reasons are described.

Table 7.2 Explanations of WTP values

Category	Type of explanation provided	Long scale (%)	Short scale (%)
E1	Question deemed in-applicable.	2.6	0.6
E2	Subject expressed difficulties in estimating WTP	6.2	9.6
E3	WTP estimate started to be based on a nominal amount	1.7	0.6
E4	WTP reflects ability to pay (affordability)	31.6	34.8
E5	WTP reflects a fair, acceptable or reasonable value	7.8	11.8
E6	WTP reflects costs of screening	7.6	7.9
E7	WTP reflects perceived benefit of screening	8.8	10.1
E8	Reported familial experience of colorectal cancer	2.0	3.4
E9	Protest expressed at the idea of payment	18.9	19.1

As is evident from the table, there are a similar proportion of explanations offered for each of the scales. The range provided in the scale therefore had very little impact on changing the actual thought processes described by the respondents when reaching a maximum WTP value.

7.3.5 WTP for FOB and FS screening

Comparison of the WTP distributions for the long-scale and short-scale sample reveal a significant difference in location. The medians for the long scale instrument were £50, whilst those for the short scale were £40 and £45, for FOB and FS, respectively. (Mann-Whitney, Z = -3.06 and –3.64, p<0.01, for FOB and FS, respectively). The WTP distributions also differed in shape (Kolmogorov-Smirnov, Z= 2.82 and 2.74, p<0.01, for FOB and FS, respectively). Both samples (FOB short-scale and long-scale) offer similar proportion of zero values therefore the difference in location of the

WTP distribution can not be attributed to a differential number of zero values offered (8.4 and 7.9 per cent for FOB short-scale and long-scale, respectively, and 9.0 and 9.4 per cent for FS short-scale and long-scale, respectively, χ^2, p>0.4 in both cases).

Examination of the spread of the data can help to explain the differences. With the long-scale sample, 18.5% and 18.3% selected WTP values (for FOB and FS, respectively) above the maximum available to those using the short-scale instrument (£100). Therefore 18% chose values with the long-scale instrument over £100 in value. None of the subjects completing the short-scale instrument wrote a value greater than £1000 (the maximum available if completing the long-scale instrument) even though more subjects used the 'open-ended option' with the short-scale (7 people) compared to the long-scale (4 people).

7.4 Range bias

To investigate range bias, the potential effects of confounding variables, such as income and socio-demographic variables have to be taken into account as these may have an effect on WTP independently. To allow for this effect, a linear regression is used with WTP as the dependant variable. Both the long-scale and the short-scale data are combined. Due to the clustering of values, the WTP variable is expressed in logarithmic form (coefficients of skew 5.2 and 4.5 for the FOB and FS positive values, respectively). The short-scale and long-scale instrument were distinguished by a dummy variable (short-scale = 1). The prior hypothesis being that you would expect the long-scale instrument to have a positive effect on the overall WTP, i.e. the dummy variable would be significant and negative. The socio-demographic characteristics for each of the respondents, including the reasons offered were entered as independent variables.

Table 7.3 presents the results from the regression models. As expected the significance and the signs on the coefficients appeal to intuition. The results show that respondents with a high income and with more formal education are more likely to provide a greater WTP value. In addition, frequent visits to the dentist, if the respondent is particularly worried about the disease and is a non-smoker, then again, they will provide a greater WTP value for the screening tests. Three dummy variables for coded reasons behind the WTP valuation also appear in the models. First, respondents who cite tangible benefits from screening as their reason offered significantly higher values. Second, those indicating any form of 'protest' against having to value or pay for the service in question offered lower valuations. And finally, those noting that they had found the valuation exercise difficult offered higher values, although, in this case, no theoretical predication as to coefficient signs appears possible. Both models for the FOB and FS tests are very similar.

The most important result, for the purpose of this study, is the direction of the sign on the coefficient for the scale variable. As expected, it has a significant negative sign showing that the subjects who answered the short-scale are more likely to have a smaller WTP amount compared to subjects who responded to the long-scale. In fact, evaluated at the means, the models imply that the use of the short-scale instrument in place of the long-scale instrument would reduce the predicted mean WTP by 25.5% (from £50.1 to £37.3 for FOB testing) and by 28.9% (from £56.1 to £40.0) for the FS testing. Reversing this, the results seems more dramatic as by implication, the use of

Table 7.3 Regression results, Ln(WTP) as dependant

	FOB			FS		
	β	T-ratio	p =	β	T-ratio	p =
Scale length (short = 1)	-0.294	-3.018	0.003	-0.341	-3.498	0.000
Household income (<£10,000 = 1)	-0.410	-5.357	0.000	-0.331	-4.192	0.000
Household income (>£30,000 = 1)	0.305	3.574	0.000	0.313	3.750	0.000
Age on leaving full time education (years)	0.012	1.950	0.051	0.012	1.979	0.048
Number of dental visits in previous 2 years	0.059	2.923	0.004	0.062	3.081	0.002
Current Smoker? (yes = 1)	-0.147	-1.836	0.067	-0.198	-2.475	0.014
Worried about colorectal cancer? (yes = 1)	0.289	4.272	0.000	0.268	3.909	0.000
Found WTP valuation difficult? (yes = 1)	0.398	3.216	0.001	0.576	4.608	0.000
Explicitly mentioned benefits of screening? (yes = 1)	0.403	3.903	0.000	0.345	3.415	0.001
'Protest' at having to pay? (yes = 1)	-0.322	-3.775	0.000	-0.277	-3.229	0.001
Constant	3.444	24.790	0.000	3.530	25.331	0.000
Adjusted R-squared	0.127			0.139		

the long scale instrument in place of the short scale instrument would inflate predicted positive WTPs by 34.2 and 40.6 per cent, for FOB and FS, respectively.

Several other formulations of models performed just as well as the one presented in table 7.3 essentially due to collinearity between the variables, i.e. reported health status and employment is strongly associated with income and could realistically replace the latter in the regression model. The model chosen for table 7.3 was done so on the basis of coefficient significance, goodness-of-fit and theoretical justification. Whatever the precise formulation of the model, the essential point is that the coefficient associated with the scale-length dummy remained negative and significantly different from zero. When evaluated at the means, these models also entailed essentially similar proportional changes in predicted WTPs to that from the table 7.3 models. As illustrated, earlier in this chapter, there is no evidence to suggest that the distribution of reasons for valuation offered differs significantly by scale length.

7.5 Overall WTP values

Table 7.4 illustrates the mean, median and inter-quartile ranges obtained from both the long-scale and the short-scale.

Table 7.4 Overall WTP values

Test	Long-scale					Short-scale				
	Mode	Mean	Inter-quartile ranges			Mode	Mean	Inter-quartile ranges		
			25	50	75			25	50	75
FOB	50	93	20	50	100	50	60	20	40	50
FS	100	91	20	50	100	50	58	20	45	60

The effect of having a shorter scale has reduced the mean by the order of £30. The impact on the median value however is not as great with the value decreasing by £10 for FOB and £5 for FS.

7.6 Grouping analysis

Up until now, the results have shown evidence of range bias however as we are dealing with payment scale data, one further test is required. The payment scale instrument only deals with discrete amounts chosen from the pre-specified scale, whereas the individuals subjective valuation will come from a continuous range of values. For example, a subject receiving a long-scale instrument might ideally have wished to choose a value of £630, but only £600 and £700 were available. With this condition therefore the subject will presumably round their ideal values to the nearest point on the provided scale. This implies that the WTP data obtained from the payment scale are actually grouped about mid-points: the scale value of £600, for example, would have been utilised by respondents with ideal values in the range £550-£650, as individuals with values outside that range would have had closer scale points available to them.

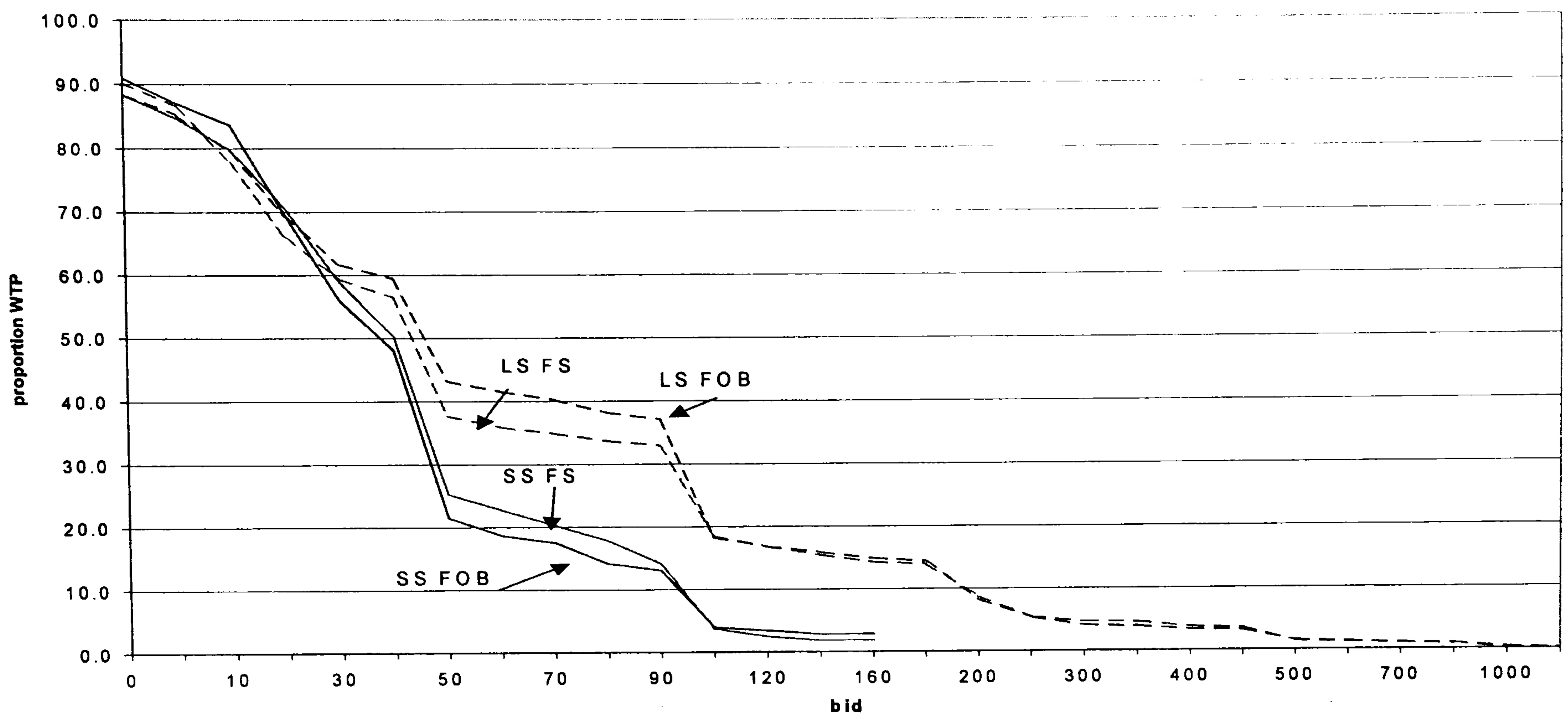
The coarseness to grouping differs between the two scales as those using the short scale would have rounded their ideal values to the nearest £5, whilst those using the upper end of the long scale would have rounded to the nearest £100. The payment scale results give no indication as to the probable distributions of ideal values about midpoints although, for our hypothesis these distributions are crucial. This is because if, in selecting their scale points, all the short-scale subjects had actually 'rounded down' from their ideal valuations, whilst all the long-scale subjects had actually 'rounded up', then the supposed range bias might be an artefact resulting from our imposition of essentially arbitrary scale points.

It is relatively easy to test the impact of the distribution on the range bias hypothesis. To test, the distribution that would have the most 'damaging' effect on the range bias hypothesis would require that all the short-scale responses were as large as possible, consistent with the observed mid-points, whilst those from the long scale were as small as possible. To construct this distribution the WTP valuations were adjusted accordingly. The short scale instrument, with uniform intervals of £5, the maximum available from the range (£2.50) was added to each of the scale values actually recorded. For the long scale, the maximum possible was deducted from each observation, for example, £10 from payment scale values between £100 and £200, and £50 from values between £500 and £1000. These values were then transformed into logarithms and the models displayed in table 7.4 were re-estimated, using the adjusted data as the dependent variables. In both cases, the effects on the models were minimal and the coefficients for the scale length dummies remained negative and significant (t-ratio = -2.838, p = 0.005 for FOB; t -ratio = -4.195, p=0.000 for FS). Even under the most extreme distribution of grouped data, therefore the range bias hypothesis appears robust.

7.7 Demand curves

Another method of presenting the data is to map out demand curves for the screening programmes. Figure 7.2 illustrates the demand curves for FOB and FS using the data obtained from the short and long scale.

Figure 7.2 WTP for colorectal cancer screening (short-scale (SS) and long-scale (LS))



Both of the curves (for short and long scale) are very similar until the bid reaches the £30 point, after that point the long scale diverges from the short scale. As one would expect at every bid (above £30) the demand curve produced from the long scale data lies above and to the right of the short scale demand curve.

7.8 Discussion

Mitchell and Carson advocate the use of the payment scale format as it provides respondents with a context of values without providing them with a specific starting point. The payment scale style of question however has its own biases, which Mitchell and Carson do recognise. The main bias inherent within the design are central point bias and range bias that result from the choice of scale provided to the respondent.

Chapter six described a large WTP study designed to elicit values for the perceived benefits from screening for colorectal cancer. This study estimated that, using the payment scale design, the WTP for these benefits is approximately £50 (median) and £90 (mean) respectively. This result was produced from applying a scale that ranged from £0 to £1000.

Table 7.5 summarises the main findings from this chapter. To test for the effect of range bias (and central point bias), a new shorter scale was administered (£0-£100) and average WTP estimated. The results obtained from the new shorter scale suggest that range bias has an impact on the overall mean results from the WTP question however the effect on the median WTP is not so pronounced. It was found that mean

Table 7.5 Summary of findings from chapter seven.

Findings
1. A similar proportion of respondents failed to respond to both the short scale (SS) and the long scale (LS) instruments (12.5% and 11.9% for the LS and SS, respectively).
2. A similar proportion answered just one WTP question (11.6% and 7.9% for the LS and SS, respectively).
3. The majority of respondents preferred FOB and there was no significant difference in preference groups between the two scales.
4. A similar proportion of qualitative explanations for the WTP values were offered for both the scales.
5. The WTP distributions between the two scales differed by shape (Kolmogrov-Smirnov, Z = 2.82 and 2.74, p<0.001).
6. The medians for the LS instrument were £50, whilst the medians for the SS instrument was £40 and £45 for FOB and FS, respectively.
7. Both the SS and the LS offered similar proportions of zero values.
8. With the LS sample, 18.5% (FOB) and 18.3% (FS) selected WTP values beyond the SS instrument (values > £100).
9. None of the respondents answering the SS instrument chose WTP values > than £1000 (maximum WTP value in LS instrument).
10. More respondents used the open-ended option at the maximum point of the SS than with the LS.
11. The direction of the sign on the coefficient of the scale variable is negative and is significant (dummy variable: short scale = 1).
12. Respondents answering the SS instrument are more likely to provide a smaller WTP amount compared to subjects who responded to the LS instrument.

WTP values produced from the short and the long scale differed by approximately £30, the median WTP only differed by a maximum of £10 (£10 for FOB and £5 for FS).

The fact that the medians produced from the short and long scale (£50 for FOB/FS for the long scale and £40 (FOB) and £45 (FS) for the short scale) are very similar would appear to counter the proposition that the payment scale format design is prone to 'central-point' bias (Herriges and Shogren 1996). Examination of the scales show that although the median WTP falls at the mid-point on the short scale, this is not the case for the long-scale. The long-scale median WTP value is the seventh value from the 'top' of scale therefore is some distance from the mid-point.

However large differences arise when the mode and mean values are analysed. The modal value for both of the scales differed by a substantial amount, for the short scale instrument it was £50 for both FOB and FS while for the long scale instrument the modal value reached as high as £100 for the FOB test.

The 'problem' of range bias within a payment scale instrument depends on what the WTP results are being used for. If the study requires precision within the results, therefore an analysis of the mean values, the use of the longer scale instrument would, in this case, have led to the reporting of means around 55 per cent higher than would have been the case using the short scale instrument. However if the analysis had been restricted to the reporting of median values the difference between the two scales would have been a maximum of £10. These results (especially when focusing on mean values) are suggestive that range bias indeed exists within a payment scale format with a longer scale leading to a higher overall average WTP value.

The purpose of this short study was to examine the affect of range bias by comparing a previous instrument implemented (the long-scale) against a new shorter scale instrument. As shown, the results are suggestive that range bias exists. As to which scale is the most appropriate to adopt when conducting a WTP study – the answer is not clear. It seems that, what Mitchell and Carson do acknowledge (but have not yet resolved) is that extending the scale so as not to discourage those with high WTP valuations is likely to over-encourage those with low valuations (Mitchell and Carson 1989). To minimise this effect would require a prior knowledge of the distribution of WTP values in order to minimise the suppression of high values against the inflation of low ones. However if this set of data is known prior to conducting the study, there would seem little concern to develop the payment scale instrument in the first place.

Methodological changes that produce varying results encourages a movement towards standardisation. However a standard scale would be very challenging to create as it is highly likely that the range of WTP values produced will differ according to health care intervention and population. Different health care populations may have different associated socio-economic characteristics and income levels that lead to a difference in WTP. In addition, individuals may find it somewhat strange to be faced with a range for their WTP for an intervention that avoids a high probability of death that only reaches £100. Alternatively, if the individuals are asked to value an intervention that avoids a trivial probability they would not expect to be faced with a range reaching values as high as £1000. Whilst it seems a logical conclusion to advocate standardisation, this seems unpractical. One may argue therefore that if the analyst wishes to prove a high contingent valuation for an intervention when using the payment scale format they can use as long a scale as they believe they can get away with

Up until now, the open-ended and payment scale question formats have been investigated. The next chapter introduces a different elicitation format, that is, the dichotomous choice framework (closed-ended method). The results of using this type of elicitation instrument are discussed and then compared to the open-ended and payment scale formats.

Chapter Eight - Closed ended study

8.1 Introduction

Despite the wealth of debate surrounding what is the most appropriate question format to adopt when eliciting WTP values, empirical work comparing the 'performance' of the different formats within the health care field is fairly limited. Cited drawbacks of the open ended question design, discussed in chapter four, suggest that respondents have an increased tendency to consider 'cost'. Also, since there is a lack of guidance provided for respondents within the design this leads to an unacceptable amount of non-responders and protestors. Critics of the payment scale design have made reference to the potential of bias caused by the choice and the position of the values within the chosen scale. Chapter six describes a study designed to examine the performance of the open-ended versus the payment scale question format when eliciting WTP values for two alternative screening programmes. The results showed that the WTP values were directly affected by the choice of question format. The payment scale design produced a median WTP in the order of £50 for Faecal Occult Blood testing (FOB) and Flexible Sigmoidoscopy (FS) compared to £30 and £50 for the open-ended design, respectively. From a methodological perspective, the open-ended design produced more non-responders, a lower WTP on average, fewer protest and zero valuations and a similar amount of respondents referring to the resource cost of screening. Therefore the results suggest that the payment scale design does perform better than the open-ended in some areas but not in others. Chapter four briefly outlined the published guidelines produced by the National Oceanic Atlantic Association (NOAA) on the use of the contingent valuation method when applied to an environmental setting. One key recommendation from these guidelines was that a closed-ended question design should be adopted when eliciting WTP values. NOAA advised the choice of this design as it resembles 'real-life' decisions that people have to make on an everyday basis therefore it mimics market behaviour where the respondent is presented with a price (bid) and the decision whether to accept or reject it. The bid levels are then varied across the study sample to calculate average WTP. As the question design is believed to resemble 'real-life' decisions it is assumed less likely that respondents will answer strategically, therefore more conservative values will be sought. WTP values for colorectal cancer screening, collected through the open-ended and payment scale formats, have already been analysed and presented in chapters 6 and 7, this chapter presents a third study administered using the closed-ended format. Comparisons of the results from all three-question formats are then made. Although the recommendations from NOAA advise that the WTP values should be elicited in an interview setting using the closed ended technique, for consistency and practical reasons it was decided to conduct the "closed-ended study" using the same study protocol as before (using a postal questionnaire).

8.2 Method

8.2.1 Questionnaire design

The closed-ended questionnaire remained exactly the same to the questionnaire distributed in the open-ended/payment scale study that is discussed in detail in chapter six and seven. Briefly, the first two pages of this questionnaire contained a description of colorectal cancer, the principle of screening and of the two screening options, FOB and

FS. The first question asks subjects if they would choose to have a colorectal cancer screening test and if so, which one would they prefer. The subjects are then asked for their WTP value for each test using the closed-ended question design. This style of question asks subjects if they think having the complete series of FOB tests/ the one-off Flexi-scope test is worth £ bid amount of which they are invited to answer 'yes' or 'no'. The closed-ended questionnaire is presented in appendix six.

After each WTP question the subjects were asked to explain their reasons for their answer (questions 4 and 6 on the questionnaire). As in the open-ended /payment scale study the qualitative information obtained from the reasons were used to understand the thought processes of the respondents when answering the WTP question. A key objective to the closed-ended study was to ascertain if this would be a different thought process to that observed in the previous open-ended/payment scale study given that the respondents were answering a different style of question. The same coding frame that was developed during the open-ended/payment scale study was applied to interpret the closed-ended WTP reasons. Each of the three researchers from the health economics team at Nottingham took a sample of reasons that were selected at random and through reading and re-reading them developed themes and categories. These themes were then re-worked within the research team until eventually there were no inconsistencies in the categorisation of explanation.

When designing a closed-ended study it is advised that a pilot study should be run to determine the distribution of the bid values to be selected (Hanemann and Kanninen

1996). In this study, the open-ended/payment scale study was regarded as the 'pilot' study hence the bid amounts were chosen on the basis of the open-ended/payment scale distribution detailed in chapter six. To explain, the median WTP from the open-ended/payment scale study came to £50. This value (£50) was selected as one offer value with four more spaced symmetrical around the median. It was therefore thought that the bids of £10, £25, £50, £100 and £200 would adequately cover the expected closed-ended distribution. The bids were located at the 13th, 37th, 69th and 86th percentile of the pilot FOB WTP distribution, and the 15th, 36th, 70th and 89th for the FS distribution. This range was also chosen to test for methodological issues such as psychometric anchoring effects that can influence response in a closed-ended framework. An extreme form of anchoring is referred to as 'yea-saying' caused by respondents saying 'yes' to any bid regardless of the amount (Mitchell and Carson 1989).

The final part of the questionnaire remained exactly as before asking subjects for routine socio-demographic and economic information such as gender, age, age on leaving full time education, employment status and household income.

8.2.2 The distribution

The closed-ended questionnaires were distributed using the Trent Focus Collaborative Research Network (CRN) of general practitioners (GPs) that were used in the initial open-ended/payment scale study. To begin, GPs were asked if they would like to participate in an extension of the original study using exactly the same protocol as before. In response to this request a sub-sample of the original set of GPs agreed to take part and distribute the closed-ended questionnaires using the same exclusion criteria that was adopted before. To reiterate, the GPs were requested to give the questionnaire to no one under the age of 25, who had recently been diagnosed with colorectal cancer or to anyone with reading, language and learning difficulties. Six primary care practices consisting of 19 GPs offered to take part. Each GP received 42 questionnaires meaning that approximately 800 closed-ended questionnaires in total were sent out for distribution.

8.2.3 Regression analysis

The response to the closed-ended WTP questions were assessed using regression analysis with the logit model. A separate model for the FOB and FS screening programme was created. The linear regression models produced in the open ended/payment scale study showed that males, household income, age on leaving full-time education, health motivation, above average perceived risk and frequency of visits to the dentist all had a positive association with WTP. It was decided therefore to replicate these linear models for the closed-ended sample using a logit regression. Our prior hypothesis was that all those variables that had significantly influenced the WTP distribution in the previous study, would have the same effect in this study. The WTP values are then analysed with respect to the reasons provided by each respondent. To illustrate the relative demand for the screening options, the proportion of subjects saying 'yes' to each bid, for each test, is then plotted.

8.3 The response

The response rate varied across the six practices that participated in the study. Practice response rate was calculated as the number of completed questionnaires returned to the university. It is not clear from the data set if the response rate was influenced by either the GP failing to offer the questionnaire or the respondent declining to complete it as unfortunately the GPs were unable to record the details of the non-responders. In total, 263 completed questionnaires were returned giving an overall study response rate of 33%. Table 8.1 displays the variation of response rate across the practices.

Table 8.1 Practice response rate

Practice	Number of questionnaires	Response rate (%)
1	87	41
2	66	52
3	52	41
4	22	52
5	33	78
6	3	1

This table shows that the response rate varied from as low as 1% in one practice to 78% in another. Again, as before in the previous study, this rate was largely influenced by the individual personalities of the GPs involved.

8.4 Socio-economic characteristics

Table 8.2 displays the socio-economic characteristics of the closed-ended sample. For completeness the data is categorised by practice and then illustrated for the whole sample.

Table 8.2 Socio-demographic characteristics

Variables	Sample characteristics						
Practices	P 1	P 2	P 3	P 4	P 5	P 6	Total
n	87	66	52	22	33	3	263
Gender:							
Female	59 (68.6%)	41 (65.1%)	34 (65.4%)	14 (63.6%)	19 (57.6%)	2 (66.7%)	169 (65.3%)
Age [median]	50	51	62	40	44	61	50
Proportion under 45 years	28 (33.3%)	23 (36.5%)	13 (25%)	13 (38.1%)	17 (81.8%)	1 (33.3%)	95 (37.1%)
Income:							
<£10K	16 (22.5%)	17 (27.9%)	13 (26%)	6 (31.6%)	6 (22.2%)	1 (33.3%)	59 (25.5%)
£10–20K	25 (35.2%)	26 (42.6%)	14 (28%)	11 (57.9%)	9 (33.3%)	2 (66.7%)	87 (37.7%)
£21–30K	10 (14.1%)	14 (23%)	5 (10%)	0 (0)	5 (18.5%)	0 (0%)	34 (14.7%)
>£30K	20 (28.2%)	4 (6.6%)	18 (36%)	2 (10.5%)	7 (25.9%)	0 (0%)	51 (22.1%)
Proportion quite /very worried about bowel cancer	21 (24.1%)	12 (18.2%)	4 (7.7%)	4 (18.2%)	6 (18.2%)	1 (33.3%)	48 (18.3%)
% who think chances of developing disease higher than average	6 (7.2%)	4 (6.3%)	4 (8.2%)	0 (0)	3 (9.7%)	1 (66.7%)	18 (7.1%)
Number of visits to dentist in last 2 years	4 [2-5]	4 [2-5]	5 {3-5]	2.5[1-5]	4[1.5-5]	6[5-6]	4[2-5]
Age on leaving full time education [median]	16	16	17	16	16	15	16
Smokers (%)	20 (23.5)	14 (21.5%)	1 [2]	7(31.8%)	9 (28.1)	0 (0%)	51 (91.8%)
Health motivation score [median]	4.25 [4-4.5]	4.25 [3.75-4.5]	4.25 [4-4.5]	4.25 [3.94-4.5]	4.25 [3.75-4.5]	4.75 [4.5-5]	4.25 [4-4.5]

Since the practice response rates were variable it was felt that it was important to make sure that each practice sample was representative of the total sample. Table 8.3 extends this analysis by categorising the data by bid level to confirm that there were equal proportions of respondents within each bid category. Overall, there are a slightly greater proportion of females (65%), the median age is 50 with 37% under the age of 45 years, 18% are quite or very worried about the disease and 7% think that their chances of developing the disease are greater than average. The greatest proportion of the sample fall into the £10-£20K income category (37%) with the least in the £21-30K category (15%).

8.5 Preferences

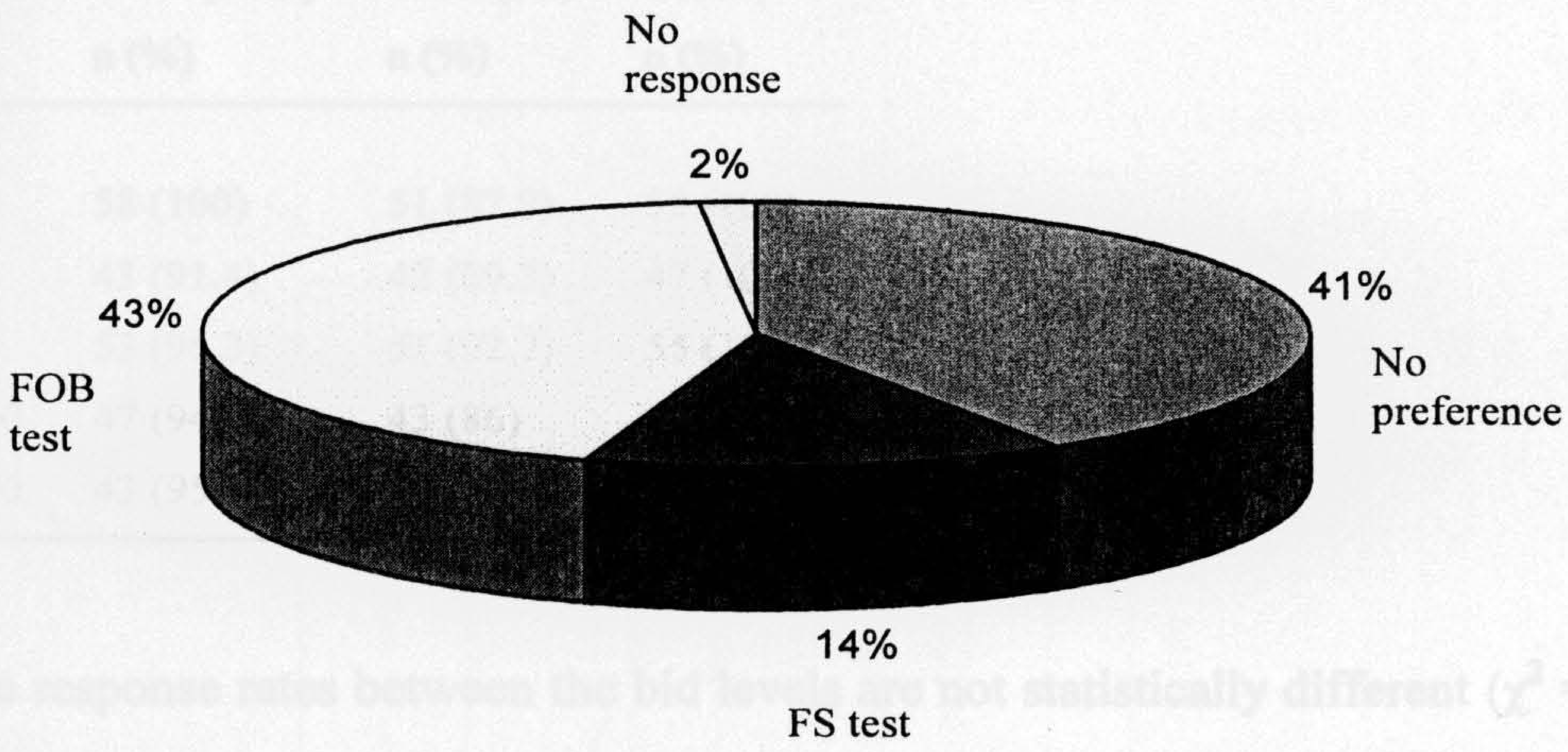
A total of eight subjects stated that they would not be willing to have a screening test. The characteristics of this sample were not statistically significant from the rest of the sample (χ^2 = .435, p=.804).

The remaining group of respondents chose the series of FOB tests as the overall preferred option as illustrated in figure 8.1.

Table 8.3 Sample characteristics by bid amount

Variables	Sample characteristics					
Bids	£10	£25	£50	£100	£200	Total
n	59	46	55	52	47	263
Gender:						
Female	39 (66.1%)	32 (69.6%)	37 (67.3%)	31 (59.6%)	30 (63.8%)	169 (65.3%)
Age [median]	53	46	47	57	49	50
Proportion under 45 years	19 (33.3)	20 (43.5%)	24 (42.9%)	13 (25.5%)	19 (41.3%)	95 (37.1%)
Income:						
<£10K	17 (31.5%)	7 (17.9%)	14 (28%)	13 (28.3%)	8 (19%)	59 (25.5%)
£10–20K	16 (29.6%)	18 (46.2%)	21 (42%)	17 (37%)	15 (35.7%)	87 (37.7%)
£21–30K	10 (18.5%)	6 (15.4%)	5 (10%)	6 (13%)	7 (16.7%)	34 (14.7%)
>£30K	11 (20.4%)	8 (20.5%)	10 (20%)	10 (21.7%)	12 (28.6%)	51 (22.1%)
Proportion quite /very worried about bowel cancer	9 (15.3%)	6 (12.8%)	16 (27.6%)	9 (17.3%)	8 (17%)	48 (18.3%)
% who think chances of developing disease higher than average	6 (10.3%)	3 (6.5%)	3 (5.4%)	3 (6.3%)	3 (6.8%)	18 (7.1%)
Number of visits to dentist in last 2 years	4[2-5]	5[3-5]	4[2-5]	3[2-5]	5[1-5]	4[2-5]
Age on leaving full time education [median]	16	16	16	16	16	16
Smokers (%)	9 (15.5%)	11 (23.9%)	7 (12.3%)	14 (28)	10 (21.3%)	51 (91.8%)
Health motivation score [median]	4.25 [4-4.5]	4.25 [3.75-4.5]	4.25 [4-4.5]	4.25 [3.75-4.5]	4.25 [3.75-4.5]	4.25 [4-4.5]

Figure 8.1 Preference groups



These preference proportions were not statistically different from the preference direction recorded in the previous open ended/payment scale study [46.3%(FOB), 13.8%(FS), 38.8%(no preference)], (χ^2 = 5.31, p = 0.26).

8.6 Willingness to pay response

As explained, advocates of the closed-ended format promote its 'easy' design and claim that respondents find it easier to understand which in turn leads to a higher response rate and a more 'conservative' WTP value. In the closed-ended sample, 86% responded to both WTP questions (FOB and FS), 7% responded to one WTP question and 7% failed to answer both WTP questions. Table 8.4 displays the response by the bid level for the group of respondents that replied 'yes' or 'don't know' to the 'would you like a screening test' question (n=255).

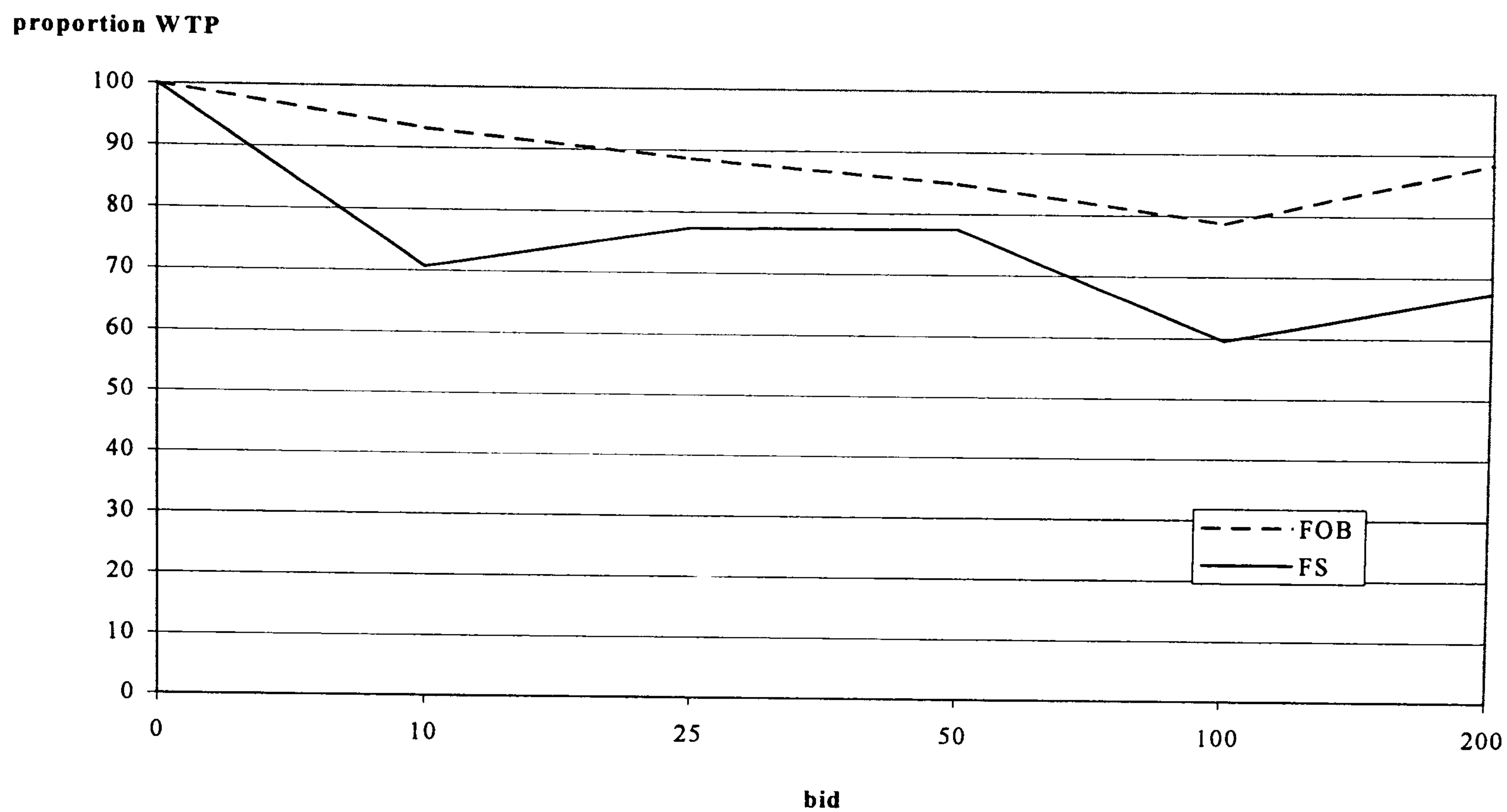
Table 8.4 Response by bid-level (n=255)

Bid	WTP (FOB) n (%)	WTP (FS) n (%)	Total n (%)
£10	58 (100)	51 (87.9)	58 (100)
£25	43 (91.4)	42 (89.3)	47 (100)
£50	53 (96.3)	51 (92.7)	55 (100)
£100	47 (94)	43 (86)	50 (100)
£200	43 (95.5)	40 (88.8)	45 (100)

The response rates between the bid levels are not statistically different (χ^2 = 3.48, p=.481 (FOB question), χ^2 = 1.55, p=.818 (FS question)) therefore the bid level offered does not affect the completion rate of the WTP question.

Of those respondents that did respond to at least one WTP question (n = 245), figure 8.2 displays the relationship between the bid level and the proportion of the sample of respondents that indicated 'yes' they are WTP. Theory would predict that as the bid level rises the proportion of respondents in our sample that are WTP would fall. This would then provide a demand curve for FOB and FS which can be used to estimate average WTP values. For the WTP to be accurately estimated, there must be a sufficient percentage of subjects saying 'no' to the highest bid (£200) (Ryan, Ratcliffe et al. 1997). Figure 8.2 shows that at the bid £200, as much as 88% (for FOB) of the sample were still agreeing to pay. This is a surprising result given that at £100 we have a lesser proportion of respondents being WTP (78% (for FOB)).

Figure 8.2 Proportion of sample WTP as function of bids (£10-£200) (n = 245).



8.7 Inclusion of higher bids

This result meant that the closed-ended bid distribution had to be modified to include bids beyond £200. This was with the intention of trying to reduce the proportion of respondents saying 'yes' at the highest bids. Consequently two higher bids were added to the bid design and new questionnaires were distributed to two new set of samples applying the bids £500 and £1000. The bid design then resembled the following scale:

£10
£25
£50
£100
£200
£500
£1000

The scale starts off at £10 then gradually increases with greater increments applied. To distribute the £500 and £1000 bid questionnaires the GPs were once again asked if they would be willing to hand out the questionnaires using the same protocol. One practice volunteered and 140 questionnaires were sent out for distribution.

8.8 Additional bid response

Ninety-one completed questionnaires were returned achieving a response rate of 65%. Given that the £500 and £1000 samples had been recruited through just one practice it was important to check the representativeness compared to the previous sample. Table 8.5 displays the sample characteristics for the £500 and £1000 bid and illustrates that the two sets of respondents are similar to the previous closed-ended sample of respondents in terms of gender and age. Both the £500 and £1000 groups have a greater number of subjects that fall into the '<£10K' income category and less subjects are in the '£30K and over' category however these differences are not statistically significant (χ^2 = 15.69, p = .015).

Table 8.5 Sample characteristics for two additional bids.

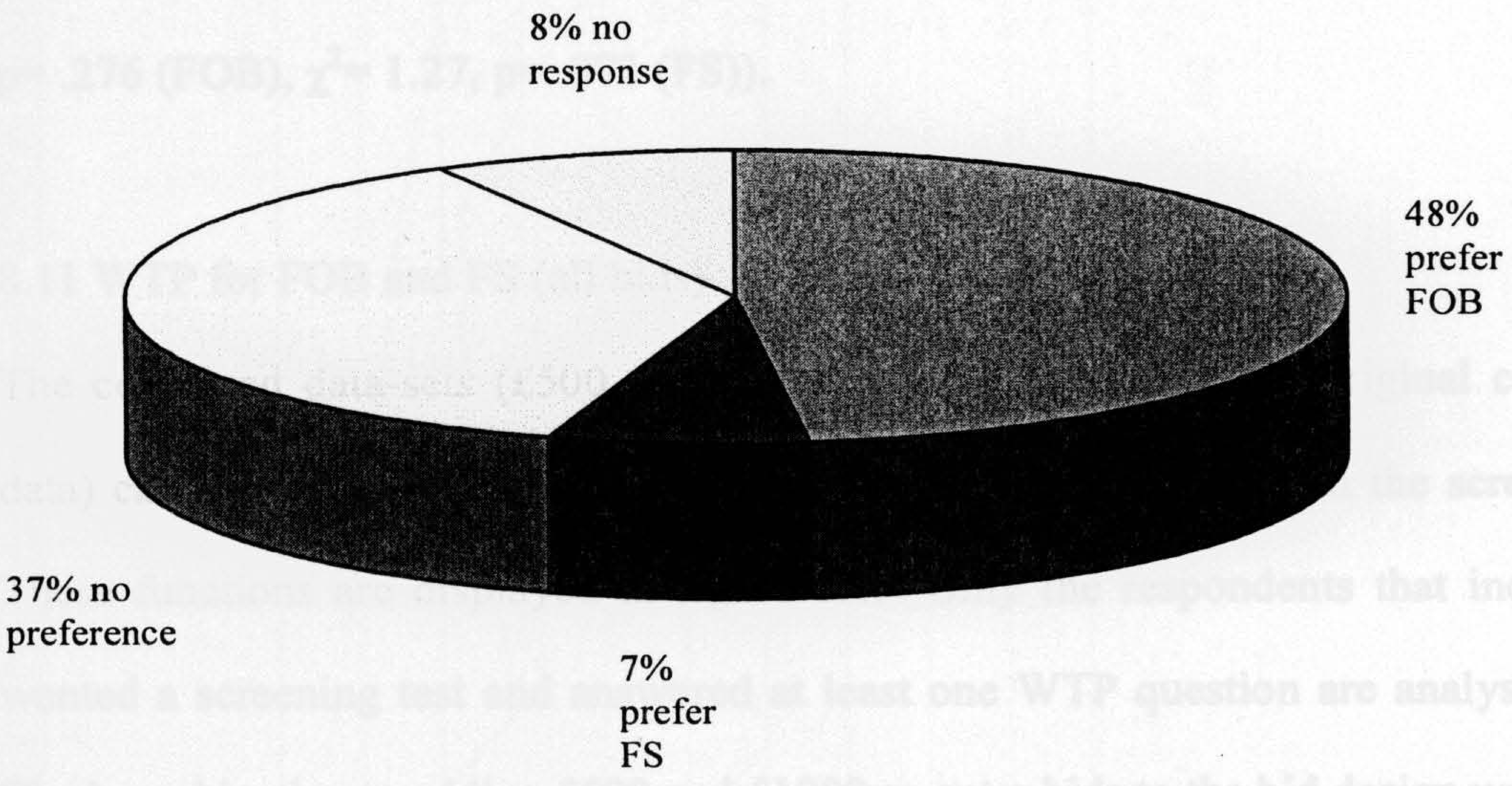
Variables			
Bids	£500	£1000	Total sample
n	44	46	354
Gender:			
Female	25 (56.8%)	29 (63%)	223 (63.9%)
Age [median]	45	54	52
Proportion under 45 years	31 (72.1%)	34 (75.6%)	118 (34.3%)
Income:			
<£10K	17 (44.7%)	17 (48.6%)	93 (30.6%)
£10–20K	12 (31.6%)	11 (31.4%)	110 (36.2%)
£21–30K	7 (18.4%)	4 (11.4%)	45 (14.8%)
>£30K	2 (5.3%)	3 (8.6%)	56 (15.8%)
Proportion quite /very worried about bowel cancer	9 (20%)	10 (21.7%)	67 (18.9%)
% who think chances of developing disease higher than average	5 (11.4%)	5 (11.6%)	28 (8.3%)
Number of visits to dentist in last 2 years	4[1-5]	4[2-5]	4[2-5]
Age on leaving full time education [median]	15	16	16
Smokers (%)	9 (20%)	13 (28.9%)	73 (21%)
Health motivation score [median]	4 [1.75-4.5	4 [2-5]	4.25 [3.75-4.5]

8.9 Additional bid preferences

Seven respondents that received either the £500 or £1000 bid indicated that they would not want a screening test. This means that judging from these responses, 15 respondents in total (8 respondents from original bids £10-£200) would decline an offer of a colorectal cancer screening test.

The preference groups for the £500 and £1000 bid sample are similar to the previous bids in that the majority prefer the FOB test. Figure 8.3 illustrates the preference groups for the £500 and the £1000 bid sample.

Figure 8.3 Preference groups for the £500 and £1000 bid sample.



When the preference groups are analysed for the full sample (all bids) the majority prefer FOB (42%), 12% prefer FS and 41% have no preference between the tests (5% no response rate).

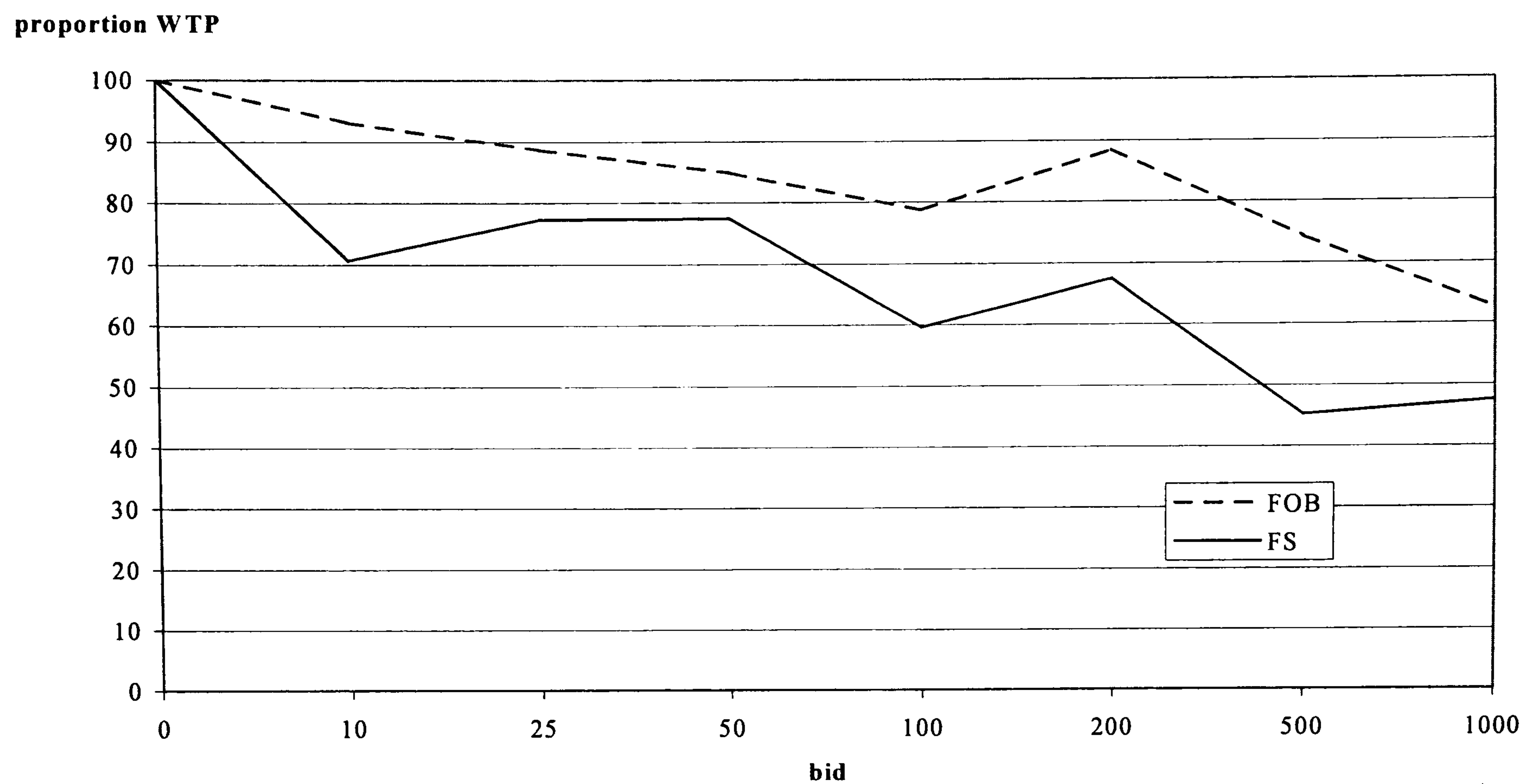
8.10 WTP response (£500 and £1000 bids)

As mentioned, seven respondents in the £500 and £1000 bid groups said 'no' to having a colorectal cancer screening test this means that 84 subjects from this group were presented with the WTP question, 40 in the £500 bid group and 44 in the £1000 bid group. To check if the response rate to the WTP question varies by bid level (as done in the £10-£200 group) the response to the £500 and £1000 is analysed. The £500 bid WTP question achieved a response rate of 77.5% for FOB and 72.5% for FS. The £1000 bid WTP question achieved a response rate of 86.3% for FOB and 79.5% for the FS test. If we refer to table 8.4 to look at the response rates to the WTP question for the other bids we can see that these response rates (for £500 and £1000) are lower. However the response rates across all seven bids for both tests are not statistically different ($\chi^2 = 7.51$, $p= .276$ (FOB), $\chi^2= 1.27$, $p= .973$ (FS)).

8.11 WTP for FOB and FS (all bids)

The combined data-sets (£500 and £1000 bid data added to the original closed-ended data) can be plotted to display the probability distributions for both the screening tests. These functions are displayed in figure 8.4. Only the respondents that indicated they wanted a screening test and answered at least one WTP question are analysed (n=314). The key objective to adding £500 and £1000 as extra bids to the bid design was to reduce

Figure 8.4 Proportion of sample WTP as function of bids (£10-£1000) (n=316).



the proportion of respondents in the sample replying 'yes' to the higher bids. The curves in figure 8.4 demonstrate that the percentage of respondents saying 'yes' has fallen with the introduction of the higher bids but the proportion is still approximately between 50 and 60% at £1000. It is intriguing that, for the FS test, the proportion falls from 67% at the £200 bid to 45% at the £500 bid but then rises again to 47% with the £1000 bid. It is difficult to ascertain the reasons why the demand curve violates monotonicity. It could be that the 'yea-saying' effect is influencing the results as individuals respond with a 'yes' to any value of bid regardless of the actual value that they place on the test. One can only guess why this occurred to a greater extent with the £1000 bid compared to the £500 bid however this result does create a cause for concern with respect to the validity of the WTP exercise. If respondents are asked to complete a university-designed questionnaire by a respected person such as a GP it may be that they regard the bid amount presented to them as one that they should be expected to pay. The curves for both tests have similar shapes corresponding to the direction of the preference groups defined earlier. As one would predict, for each bid, the proportion of subjects that are WTP the bid amount is higher for the FOB test compared to the FS test (42% prefer FOB compared to 12% that prefer FS).

8.12 Inconsistencies

The WTP methodology relies on a subject being consistent in their response to the series of questions within the questionnaire. Put another way, the theory of rationality would predict that subjects will have a higher WTP value for their preferred test over their least preferred test. As discussed in chapter four, behavourial analysts, experimental and environmental economists have investigated the frequently encountered phenomenon of preference reversal. When a respondent provides a larger WTP value for their least preferred test then they are showing signs of preference reversal while completing the questionnaire. Logic predicts that this respondent has misunderstood the task and should be eliminated from further analysis however experimentalists and psychologists argue that 'preference reversal' is essentially a normal human characteristic. It is an interesting issue within WTP literature as analysts face the question of how to deal with subjects who have shown signs of preference reversal in their answers to the study questions. In this closed-ended study we investigate the 'inconsistencies' by examining the response to each of the WTP questions by the preference groups for each of test. Table 8.6 displays the results.

Table 8.6 Response to WTP question by preference group (n=314, (2 respondents did not indicate preference group)).

	FOB		FS		Total
	Are you WTP?		Are you WTP?		
Preference Groups	Yes	No	Yes	No	
FOB	113	24	72	49	138
	81.9%	17.4%	52.2%	35.5%	
FS	27	11	27	11	39
	69%	28%	69.2%	28.2%	
No Preference	120	16	105	27	137
	87.6%	11.7%	76.6%	18.2%	

Concentrating on the FOB group first, the majority say 'yes' they are WTP for FOB (81.9%) but the majority (although a lesser amount) also say 'yes' they are WTP for FS (52.2%). The FS preference group are slightly different as exactly the same proportion respond 'yes' for both tests (69%). Finally, the third preference group, i.e. those with no preference, have more people saying 'yes' to the FOB tests (87.6% (FOB) v's 76.6% (FS)).

Testing for inconsistency in a closed-ended WTP sample is limited. With an open-ended or a payment scale study individual person's WTP are aggregated to produce average population estimate. However with a closed-ended WTP study, the individual's WTP response contributes to the average valuation of a group. There is no way of knowing if that individual would have accepted a higher WTP amount if offered. Therefore if the subject chose FOB as their preferred test and accepted both offered bid amounts for FOB and for FS, there is no way of knowing if they would have accepted a higher bid amount for the FOB test if offered. It proves impossible under the WTP closed-ended format to exhaustively test for inconsistencies to the level that can be done with open-ended and payment scale studies (Ryan and Miguel 2000).

8.13 Explanations for response

After each WTP question, the respondents were asked to provide reasons for their response. These reasons can prove invaluable in translating the subjects' response. Indeed, the reasons that the subjects provided with the open-ended/payment scale study helped to categorise the subjects into groups, i.e. protestors, those who are being inconsistent, subjects who refer to cost. There was a slight variation in protocol with the closed-ended study compared to the open/payment scale study. In the previous study subjects were requested to provide a reason for their FOB and FS decision collectively for FOB and FS, in this closed-ended study they were asked to provide reasons separately for FOB and then for FS. The questionnaire was modified in such a way so as to encourage the respondents to consider the two tests separately rather than collectively, i.e. to minimise the embedding effect. Table 8.7 displays the reasons which subjects reported as being relevant to reaching their decision.

Table 8.7 Reasons for WTP response (n=316)

Type of explanation	FOB n (%)	FS n (%)
Proportion of sample giving one or more reasons	68.4	55.1
n =	242	195
E1 Question deemed inapplicable	0.4	1.5
E2 Subject expressed difficulties in estimating WTP	5.4	7.7
E3 WTP estimate stated to be based on a nominal amount	0.4	0.5
E4 WTP reflects ability to pay (affordability)	16.9	18.5
E5 WTP reflects a fair, acceptable or reasonable value	4.1	3.6
E6 WTP reflects costs of screening	10.7	13.3
E7 WTP reflects perceived benefit of screening	62.8	56.9
E8 Reported familial experience of CRC	6.6	3.1
E9 Protest expressed at the idea of payment	10.3	14.4

You can see from the table that the proportion of respondents citing reasons for each test are very similar. Very few respondents cite difficulty in responding as the reason for their WTP value. It is somewhat surprising that so few have referred to their ability to pay as a reason why they responded to the WTP question. Only 10.3% (FOB) and 14.4% (FS) have given protest answers. The majority of reasons cited refer to the benefits of screening. All subjects who cite benefits as their reason for responding to the WTP question, accepted the bid value.

8.14 Regression analysis

The linear models in the open-ended and payment scale study suggested that males, household income, age on leaving full time education, health motivation, above average perceived risk and frequency of visits to the dentist all have a positive impact on the WTP for FOB and for FS. These significant variables from this study were regarded as good predictors of the socio-economic characteristics that would have an influence on the likelihood of accepting bids in the closed-ended study hence the linear models were replicated for the closed-ended sample in a logistic regression. Table 8.8 represents the results.

Although most signs are in accordance with the previous study, only a minority of the coefficients are significant. As can be seen from table 8.8 the likelihood of accepting a bid amount is increased if the subject is a frequent visitor to the dentist, has a high health motivation and perceives their risk of developing cancer as being above average. As expected if the subject lives in a household that has a high income then they are more likely to accept an offer compared to someone from a low household income group. Despite the previous open-ended and payment scale model suggesting that males and those with more years of education have a higher WTP, the closed-ended model indicates that females are more likely to accept an offer at any given level, whilst the more educated are less likely.

An important result is the effect of the 'offer value' variable. The odds ratio with respect to this variable tells us that although it is statistically significant and with the correct sign the impact it is having on the acceptance is very small. The odds ratio is very close to unity meaning that the likelihood of not accepting an offer increases only marginally with increasing bid amounts.

Table 8.8 Logistic regression results

	FOB				FS			
	β	se	p=	exp(β)	β	se	p=	exp(β)
Offer value (£)	-0.001	0.000	0.007	0.999	-0.002	0.000	0.000	0.998
Gender (female=1)	0.472	0.324	0.145	1.603	0.155	0.287	0.590	1.167
Age on leaving full time education (years)	-0.039	0.028	0.170	0.962	-0.033	0.028	0.238	0.968
Household income, <£10,000	-0.224	0.364	0.538	0.799	-0.025	0.343	0.942	0.976
Household income, £20,000-30,000	0.979	0.598	0.102	2.662	0.076	0.418	0.857	1.079
Household income, >£30,000	0.629	0.501	0.209	1.876	0.070	0.402	0.861	1.073
Dental visits in previous 2 years	0.029	0.091	0.748	1.030	0.143	0.081	0.075	1.154
Health motivation score	0.718	0.263	0.006	2.051	0.864	0.258	0.001	2.373
Chances of cancer above average?(yes=1)	2.031	1.018	0.060	7.623	2.331	0.835	0.005	10.291
Constant	-1.199	1.127	0.287		-2.678	1.128	0.018	
Pseudo r2	0.20				0.25			
Percentage correctly specified	80.97				71.63			

8.15 Average WTP values for FOB and FS

To estimate WTP for the closed-ended data, a computationally-simple, non-parametric method was used. This requires the calculation of the proportion of acceptances at each offer amount. When the proportion increases with a higher offer amount, the mean proportion across the offer levels is estimated until a non-increasing proportion of acceptances is obtained. The relation between proportion of acceptances and offer amount is then modeled as a linear interpolation (regression). In such a formulation, median WTP is the value where the proportion of acceptances equals 50 per cent and 95 per cent confidence intervals can be obtained from the confidence band estimates of the interpolation. For the CE data, the median WTP is estimated to be £946.1 (CI 795.2 to 1,112.4) for FOB and £1,009.4 (CI 588.0 to 1,580.0) for FS. For a linear interpolation, median WTP is also mean WTP.

8.16 Discussion

The objective of this chapter was to describe a study conducted that adopts the closed-ended WTP question design when eliciting values for colorectal cancer screening. The environmental guidelines produced by NOAA promote the use of the closed-ended design administered in an interview setting. To be consistent with these guidelines, the closed-ended study should have been conducted using interviews and not a mail survey. However it is interesting nonetheless to assess the 'performance' of the technique when used in a mail survey and this enables a more consistent comparison with the earlier open-ended /payment scale study that used a mail survey.

The most obvious outcome from the closed-ended study is the level of magnitude of the WTP values revealed for both FOB and the FS test. The question design is recommended due to its similarity with real-life decisions that is meant to lead to more 'conservative' values being elicited. In this closed-ended study the revealed WTP amounts for FOB and FS are far higher compared to the previous open-ended/payment scale study. This result has been noted in previous health care studies (Johnson, Bergenzer et al. 1990). The closed-ended question leading to far higher WTP values may be down to 'yea-saying' inherent in the closed-ended format however it is difficult to test for this empirically.

Consistently, the FOB test received a greater response compared to the FS test. This may be partly caused by a potential ordering effect; perhaps the results would have changed if the WTP question for the FS test had been placed before the FOB question in the questionnaire.

Previous studies conducted in health care have shown that the closed-ended format does yield higher average response rates (Smith 2000). For both screening tests, in the closed-ended study, the response rate to the WTP question was high and certainly when explaining their reasons for their WTP response very few respondents cited difficulties in providing a response. A good proportion of respondents referred to the perceived 'benefits' of the screening test suggesting that they have thought about the actual 'value' of both the tests. Fewer respondents (compared to the previous open-ended/payment scale study) made comments relating to 'affordability' and 'protest' issues.

The closed-ended study has highlighted the disparity in results between those produced by this study and the previous open-ended/payment scale study. The closed-ended format has produced the higher mean and median WTP values but even at the highest bid offered, a substantial proportion of the sample were still saying 'yes' that they would be WTP. The most challenging aspect of the closed-ended study was the selection of the bids prior to the study being conducted. The bid design was based on the WTP distribution obtained from the open-ended/payment scale study but it was quickly realised that this range was not adequate and higher bids had to be issued. This process has happened before in closed-ended studies conducted in health care (Ryan, Ratcliffe et al. 1997).

Table 8.9 summarises the findings from this chapter. This study was conducted in a relatively controlled context with exactly the same questionnaire being administered and the only variation being a change in the format of the WTP question. The results do have important practical implications as they suggest that the actual valuation is not independent of the valuation procedure. It seems that the choice of a closed-ended question design produces a higher mean and median WTP value compared to the open-ended or payment scale design.

Table 8.9 Summary of findings from chapter eight

Findings
1. The original closed-ended bid distribution had to be modified to include bids greater than £200 due to the amount of respondents saying 'yes' to the highest bids.
2. The majority of respondents prefer the FOB test (42%(FOB), 12%(FS), 41%(no preference).
3. The response rates to the WTP question across all seven bids are not statistically different (χ^2=7.51, p=2.76 (FOB), χ^2=1.27, p=0.973 (FS)).
4. By adding the two higher bids (£500 and £1000) the proportion of respondents saying 'yes' decreased but only to just below 50%.
5. The proportion of respondents saying 'yes' produce curves that have similar shapes for both tests with the proportion saying 'yes' being slightly higher for the FOB test. This result corresponds with the direction of preferences.
6. 68.4%(FOB) and 55.1%(FS) of the sample provided qualitative reasons for their response.
7. Very few respondents cited difficulty in responding and ability to pay as the reason for their response to the WTP question.
8. Only 10.3%(FOB) and 14.4%(FS) gave protest reasons.
9. The majority of reasons referred to the benefits of the tests.
10. The effect of the 'offer value' variable in the regression was very small meaning that the likelihood of not accepting an offer increases only marginally with increasing bid amounts.
11. Median WTP is estimated to be £946 for FOB and £1009 for FS using a computationally simple non-parametric method.
12. The closed-ended elicitation format estimated WTP values that were greater than the open-ended and payment scale format – this may be down to a 'yea-saying' effect.

However the debate does not finish here, the choice of elicitation design does not just focus around open-ended versus payment scale versus closed-ended. There remains a fourth method of eliciting WTP values that a research study can adopt when conducting a contingent valuation study. This method is called the bidding game and will be the subject of chapter nine.

Chapter Nine – Iterative Bidding Study

9.1 Introduction

Chapters six to eight have presented three studies designed to estimate willingness to pay (WTP) for colorectal cancer (CRC) screening. These studies adopted the use of the open-ended (OP), payment scale (PS) and the closed-ended (CE) elicitation techniques administered using self-completed questionnaires that were returned through the post.

The study presented in this chapter adopts a slightly different approach to estimating WTP for CRC screening. As before, WTP is elicited for two alternative screening programmes, the faecal occult blood (FOB) test and the flexible sigmoidoscopy (FS) test, only this time the WTP is measured using the iterative bidding (IB) technique in an interview setting. What has been learned from conducting the WTP studies presented in chapter six to eight is the value of the qualitative information that is attached to the WTP data. The depth of the qualitative information can help to provide insight into the understanding of the WTP methodology and can also either validate or invalidate the WTP results. The hypothetical nature of the WTP method, applied in a health care setting, makes it difficult to validate respondents' behaviour, i.e. cannot compare hypothetical to real values. Consequently, the study format has to be designed to verify that the respondents are understanding and completing the WTP task in the manner that is expected. Designing the study so that the WTP values are elicited by interviews provides an opportunity to collect a richer data set (compared to mail questionnaires) and allows a much deeper analysis of the 'thought processes' of the individuals when estimating their WTP for CRC screening. The objective of this study therefore is to firstly, estimate WTP using the IB format and make general comparisons with the alternative elicitation formats that have already been presented, and secondly, assess the value of estimating WTP using interviews as opposed to mail questionnaires. Since there is a fundamental change in study protocol between this study and the previous studies (interviews instead of mail questionnaires), the WTP values cannot be directly compared. However the process and the perceived understanding of the WTP task can be compared with some general conclusions about the biases inherent within the IB method compared to the alternative elicitation methods.

9.2 Recruitment of respondents

The study was administered through a General Practice Surgery situated in Nottingham, England. Potential respondents were randomly selected from a list of patients registered with this practice. After attaining ethical approval, 1000 letters were posted to a random sample of patients who met the exclusion criteria applied to the three studies presented in chapters six to eight. These were to exclude anyone under the age of twenty-five years, with reading, learning or language difficulties and to anyone who had recently been diagnosed with colorectal cancer. The letter was printed on practice-headed notepaper and contained a short description of the study with an invitation to participate by giving up fifteen minutes of time to go into the practice to answer a few questions. This letter is presented in appendix seven. The invitation letter contained an attached slip that the respondent could complete if interested and a pre-paid envelope to return the slip by post. All interested participants then received a second letter thanking them for their interest and a request to indicate on an enclosed timetable the days (and sessions) that would be convenient.

This second letter is presented in appendix eight. Once these timetables were received, interviews were arranged using an appointment system with the participants booked into fifteen-minute slots.

9.3 Interview design.

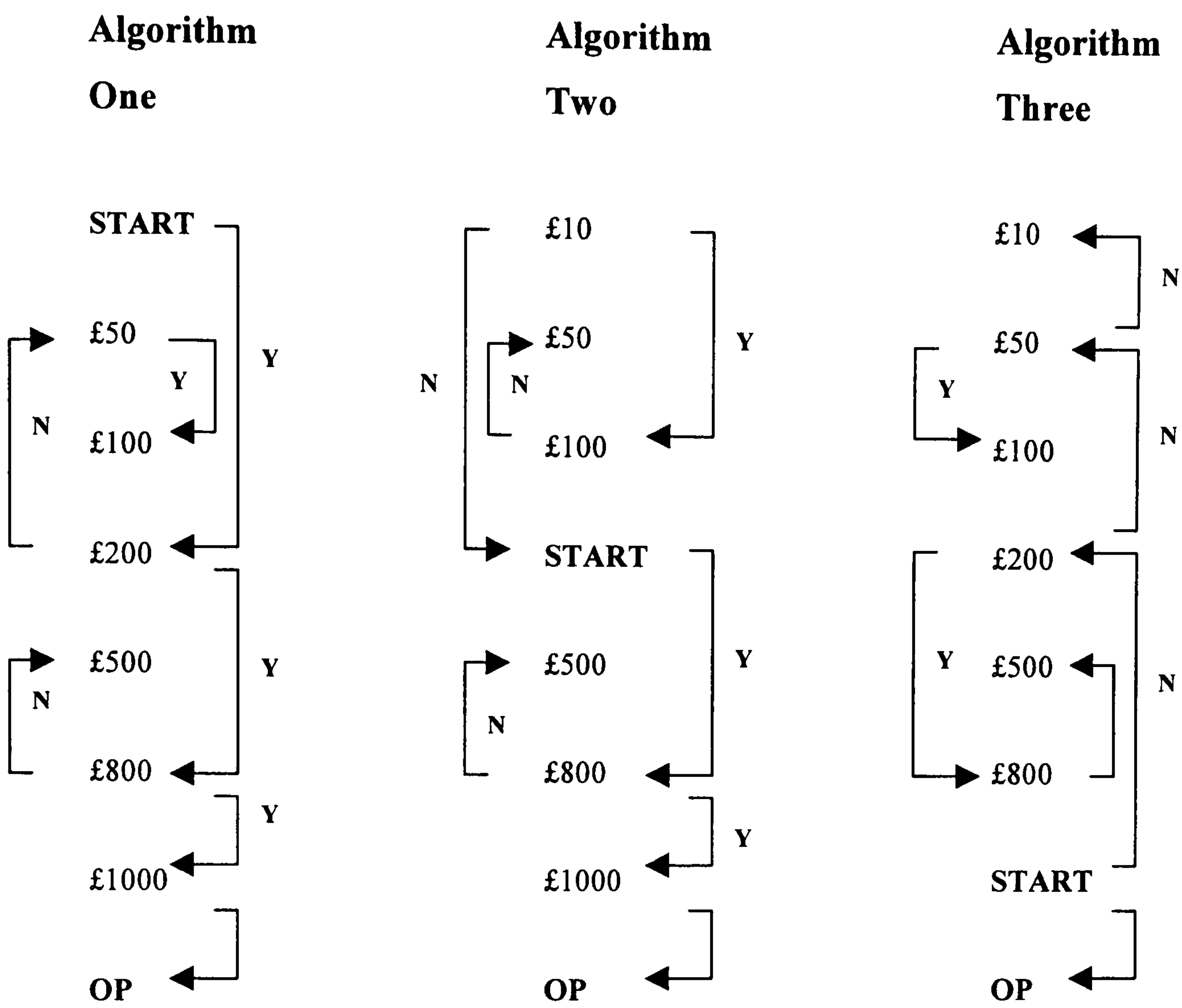
Adopting the IB technique to estimate WTP requires a lot more interaction between the interviewer and respondent compared to the previous elicitation formats. With the IB design, the respondent actively takes part in a bidding process to estimate the maximum WTP. The interviewer suggests a starting amount (bid) to the respondent and then depending on their response either raises or lowers the bid until the maximum is reached. As discussed in chapter four, the final maximum WTP value may be directly influenced by the starting bid (starting point bias) – to assess this, the respondents were randomly assigned to one of three starting amounts; £10, £200 and £1000. These starting amounts were chosen based on the WTP distribution estimated in the OP/PS study presented in chapter six and the CE study presented in chapter eight. It was expected that these starting bids would encompass the expected WTP distribution to be estimated in this study.

Before answering any questions, each respondent was provided with a short description about the study and about the nature of colorectal cancer. Explanations about the two screening programmes, FOB and FS, were then given. The descriptions of the screening programmes were kept exactly the same to the descriptions used in the previous studies, i.e. no efficacy data on the sensitivity or specificity of each test was given. This was done for two reasons, firstly, although the WTP values from this study will be administered through interviews, we wanted to keep the product that was being valued the same so no 'extra' information could be added. Secondly, it is believed that this will be the information that will be received by individuals if invited to participate in a National Screening Programme hence the WTP for each screening programme can be used as a good indicator of future screening compliance for each programme. The only difference between the studies is in the manner with which the information is communicated. Visual aids are employed in the interview to help the respondents understand the nature of the tests. If the respondent had any queries regarding either of the tests or the process of screening then these were clarified before proceeding to the next part of the interview.

As before in the previous studies the first question asked respondents if they would be willing to have a CRC screening test, and if so, the second question asked which one would they prefer, FOB, FS or no preference.

If the respondent was willing to have a screening test they were then asked for their WTP for the FOB programme and then the FS programme. As explained, three starting bids were used hence three algorithms were designed to elicit WTP of which respondents were randomly allocated to one. The use of structured algorithms has been applied before (O'Brien, Goeree et al. 1998) and help to keep the bidding process exactly the same for each respondent, as bids are either raised or lowered by the same amounts according to the respondents choice. The structured algorithms, used in the study are presented in figure 9.1.

Figure 9.1- Illustration of algorithms



For example, with algorithm one the bidding starts at £10 if the respondent responds with a 'yes' they are then asked if they would be WTP £200, if 'no', then £50, until eventually the maximum WTP is reached. Above £1000 the question design defaulted to an OP question design.

Once an agreed bid had been reached for both the FOB and the FS test, the respondents were asked for detailed reasons as to why they had chosen their preferred test. Respondents were then allowed to talk freely to the interviewer, for a short time, about their feelings towards each test. The intention behind this question was to gain further insight into the thought process of the individual and to try and understand the reasons why they preferred one test to the other. Following this, to try to gain an understanding of how well the WTP task had been understood, the respondents were asked to put a 'cross' on a 5-point horizontal scale laid out on paper to indicate how easy or hard they had found the WTP task (1-easy, 5-hard).

A major challenge of using a hypothetical WTP exercise to measure the value of a programme is the verification that the respondent has fully understood the task and answered in a reliable manner. Within the UK tax-financed health care system it is impossible to observe real-life payments for the service, if and when it is introduced. Therefore, within the interviews, to encourage respondents to consider their WTP in the context of real-life behaviour, the respondents were asked to think of a product or a service that was worth what they had just agreed they would be WTP for FOB and FS. Once the respondent had thought of a good they were then asked the WTP question again only this time the product was put in place of the WTP amount. For example, if the respondent had agreed to pay £200 for the FOB test and had recently spent £200 on a leisure break, they were asked if they would be willing to sacrifice the leisure break for the series of FOB tests. It was thought that only at this point, the respondents would realise the value (in terms of opportunity cost) of £200. If the individuals were not willing to sacrifice the product for each test, then the WTP question using the bidding process was conducted again until a revised WTP that they were willing to sacrifice was reached.

Thereafter, each subject was invited to supply us with the following sets of data that was exactly the same as the previous studies. Socio-demographic data such as gender, age, ethnic origin, marital status and age on leaving education. Economic data such as occupation, employment status, household income (in four bands, starting at zero, band-width £10,000, ending £30,000 and above). Medical history (binary variables, for both self and family) relating to experience of stomach problems, piles or haemorrhoids, heart disease, cancer, stroke and depression. Perceived own health status (four-point scale – poor, fair, good, excellent); current smoking status. They were also asked for number of visits to GP and dentist in the past one year and two years, respectively; screening record (anything other than colorectal cancer) over the previous five years and any previous experience of the FOB and/or the FS test (if any). The subjects were asked if they had any particular worry about crc (four-point scale – not at all, a bit, quite, very) and what their perceived chances were of eventually suffering the condition, compared with men and women of the subject's own age (five –point scale – much lower, lower, same, higher, much higher). Finally the subjects were asked to indicate the importance they attach to a fruit rich diet, regular exercise, breast screening and cervical screening in maintaining good health (five point scale in each of the four cases, not at all, somewhat, moderately, very, extremely). These data were coded 1 through to 5 respectively, and a mean score across the four dimensions was calculated. This score was then taken as the 'health motivation' score, i.e. the degree of belief in the efficacy of established health promotion measures, a higher mean value implies a higher motivation.

9.4 Data analysis.

To begin, the socio-demographics of the sample of respondents interviewed will be compared across the three algorithms. Overall preferences will then be measured with respect to FOB and the FS programme relative to the socio-demographic characteristics of the respondents.

The WTP for each programme will then be estimated with particular emphasis on the difference in overall WTP by algorithm, the relative difference between the WTP for FOB and FS and the reasons stated for the WTP values. This will then be followed by an evaluation of the 'opportunity cost' question and how that affects overall WTP. A significant advantage of using interviews is the depth of qualitative information obtained, this will be analysed with respect to the WTP for each screening programme.

9.5 Results

9.5.1 Response rate

One-thousand letters were posted to a random sample of patients taken from a list provided by the surgery. Attached slips, indicating a willingness to participate were received from 212 respondents, a 21% response rate overall. Time and resource constraints meant that only approximately half of these willing respondents could be interviewed. All 212 respondents received a 'second letter' and appointments for interviews were arranged on a 'first come, first served' basis. A total of 106 interviews were organised and conducted.

9.5.2 Socio-demographic data

Table 9.1 summarises the socio-demographic details for the study sample categorised by each algorithm.

Table 9.1: Socio-demographic details for the algorithm groups

	Algorithm 1	Algorithm 2	Algorithm 3	Chi-square		ANOVA	
N	33	36	34	χ^2	p	F	p
Age (mean)	54	57	56			.734	.482
Female (%)	49	52	50	.132	.936		
Educ age (mean)	17	15	16			2.666	.074
Health Status (%)				11.460	.075		
Excellent	12	8	27				
Good	70	67	35				
Fair	15	19	26				
Poor	3	6	12				
Household income (%)				4.965	.548		
<10K	14	22	18				
10-20K	29	44	43				
20-30K	39	15	25				
>30K	18	19	14				
Married (%)	76	67	79	1.566	.457		
Employed (%)	59	49	52	.259	.879		
Smoker (%)	12	17	9	.986	.611		
Worried (%)	73	58	65	4.009	.135		
Chances (%)	6	11	15	1.323	.516		
Health Motivation (mean)	4.5	4.3	4.3			1.906	.104
Illness Contacts	3.1	3.5	3.7			.970	.383

Overall, approximately 50% of the sample are female and the mean age is around 55 years. Algorithm one has slightly more respondents in the 20-30K bracket and less in the 10-20K bracket compared to algorithms two and three. There was no statistical difference in socio-demographic details between the respondents in each algorithm.

9.5.3 Preference for screening test

An overwhelming majority of the study sample indicated that they would have a CRC screening test (92% 'yes', 2% 'no'). The reasons why the six individuals chose not to have a CRC screening test are outlined in table 9.2

Table 9.2 Reasons for not wanting CRC screening

Sex	Age	Reason
Female	73	Had enough of hospitals, do not want anymore test done for rest of life.
Female	65	Don't fancy handling own stools and would not go to hospital to have a test like this one done.
Male	62	I would struggle to come forward if thought I had bowel problems – think my mates would be the same.
Male	79	I'm nearly 80 years old so I don't see the point in these tests. The country's finished anyway – everyone is drugged up to the eyeballs!
Male	38	It it's not broken, then why fix it?

The majority of the remaining sample (100 subjects) opted for the FS test over the FOB test (46%) and 19% had no preference between the tests. Only one man was unable to give an answer to this question as he 'could not make up his mind'.

A backward stepwise logistic regression can predict the socio-demographic variables that influence the preference for either screening test. This was done using a binary variable as the dependant variable (FOB preferred = 1), the results are presented in table 9.3.

Table 9.3 Logistic regression predicting preferences (FOB preferred = 1)

	β	SE	p =	exp (β)
Gender (female=1)	7.74	3.98	0.05	2,297.72
Household income (<10,000=1)	1.62	0.73	0.03	5.06
Age (years)	-0.08	0.03	0.01	0.92
Worried about colorectal cancer? (quite or very=1)	-1.36	0.72	0.06	0.26
Age leaving full-time education (females only)	-0.51	0.25	0.04	0.60
Experience of sigmoidoscopy? (yes=1)	1.32	0.68	0.05	3.74
Constant	3.87	1.77	0.03	
Pseudo – r^2	0.25			

It seems that female gender, low household income and being quite or very worried about colorectal cancer are positive predictors of a preference for FOB and absolute age, age on leaving full time education are negative predictors. Interestingly, a preference for FOB was more likely to be exhibited amongst those with previous experience of sigmoidoscopy.

9.5.4 WTP for FOB and FS.

The bidding game technique produces a range of WTP values where the individual's maximum WTP will lie. For example, with algorithm one, if the individual said 'yes' they would be WTP £10 and 'no' to £200 and then 'no' to £50 then the analyst knows that the maximum WTP lies somewhere between £10 and £49. One way of analysing this is to take the midpoint between £10 and £50 as the maximum WTP, i.e. £30. An alternative way of exploring the data is to take the lower bound (£10) and the upper bound (£49) and present the mean WTP as a range that represents the mean lower WTP, the mean mid-WTP and the mean upper WTP. The results are illustrated in table 9.4.

Table 9.4: Mean WTP by bid algorithm.

	FOB test			FS test		
	Lower	Middle	Upper	Lower	Middle	Upper
Algorithm 1	330.67	418.67	505.77	419.09	499.70	579.48
Algorithm 2	606.77	676.77	746.13	640.00	724.31	807.93
Algorithm 3	817.50	865.78	913.69	798.71	851.77	904.42

It is interesting that algorithms one and two estimate WTP in the manner that one would predict given the underlying preferences for FS over FOB and in all cases, the WTP for FS is greater than the WTP for FOB. However algorithm three has produced a different set of results with the WTP for FOB being greater than FS. Testing the preference direction within each algorithm will provide some further insight. Table 9.5 presents the results.

Table 9.5 Preference group for each algorithm

	n	Preference groups (%)			Total
		FOB	FS	No preference	
Algorithm 1	33	21.2	63.6	15.2	100
Algorithm 2	33	42.4	45.5	12.1	100
Algorithm 3	33	42.4	27.3	30.3	100
missing	1				

The results from table 9.5 highlight why it is important to check the direction of preference groups within each algorithm. This is because at first glance, given the overall preference for FS, one might have drawn the conclusion that algorithm three was producing inconsistent WTP results (a higher WTP for the least preferred test). However, when the preference groups are analysed within each algorithm it is clear that in actual fact within the group of respondents that were categorised into algorithm three, the majority of them prefer FOB hence you would expect a greater WTP for FOB over FS. These WTP results are therefore not inconsistent.

9.5.5 Ability to pay (ATP)

To ensure that the ATP of the respondents are not distorting results we need to look at the proportions of 'rich' and 'poor' within the preference groups. If there are a far greater proportion of 'rich' in the FS preference group for example, then it is not fair to compare between the WTP for FOB and FS without adjusting for this. Table 9.6 presents the results.

Table 9.6 ATP by preference group

	Preference groups n (%)		
	FOB	FS	No preference
'rich'	11 (37.9)	21 (56.8)	6 (33.3)
'poor'	18 (62.1)	16 (43.2)	12 (66.7)
χ^2	1.690	.676	2.000
p	.194	.411	.157

'rich' = £20-30K and >£30K, 'poor' = <£10k and £10-20K.

Fortunately, as table 9.6 shows there are no statistical differences between the 'rich' and the 'poor' within either of the preference groups. The sample numbers become too small to enable us to conduct any meaningful statistical tests on the proportions of 'rich' and 'poor' within each algorithm for each preference group. The overall WTP though for FOB and FS is not distorted by the difference in the ATP.

9.5.6 Starting point bias

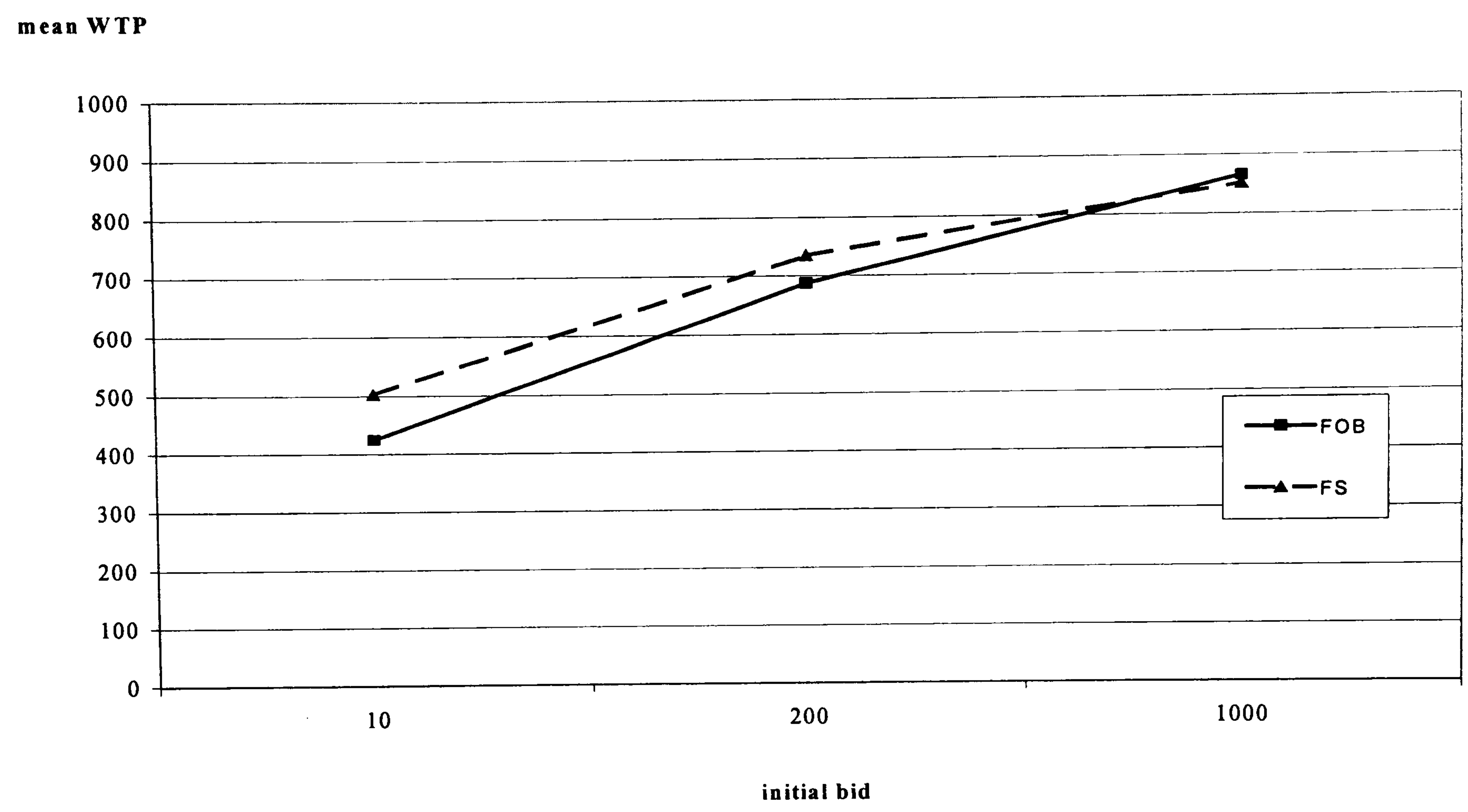
The mean and median WTP values alongside the 95% confidence intervals, for each algorithm, is presented in table 9.7. One of the key objectives to this study was to examine the effect of using different starting points, an ANOVA test is also presented in table 9.7 to examine this influence.

Table 9.7 Investigation of starting point bias

	FOB (£)				FS (£)			
	Median	Mean	95% CI		Median	Mean	95% CI	
Alg 1	350	418.67	298	541	350	499.70	389	649
Alg 2	650	676.66	525	833	650	724.31	540	883
Alg 3	1000	865.78	723	1008	1000	851.77	601	772
F	10.388				F	4.750		
P value	.000				P value	.011		

The results from the ANOVA test show that there is a statistical difference between the mean WTP between the algorithms, both for FOB and for FS. This suggests the presence of starting point bias. This data is presented graphically in figure 9.2.

Figure 9.2 Relationship between algorithm and mean WTP value



9.5.7 Regression analysis

To attempt to identify the factors that are influencing the WTP distribution an Ordinary Least Squares (OLS) regression is conducted with the mid-point WTP values for FOB and FS defined as the dependant variable. Table 9.8 presents the results from the regression where all the variables have initially entered the model and then been either accepted or rejected according to a backward stepwise procedure. This stepwise procedure terminates with the selection of the best fitting model when no additional variables can be added or deleted from the last model fitted.

Table 9.8 OLS regression coefficients (*p* values within the parentheses)

	MODEL	
Variables	WTP (FOB)	WTP (FS)
Constant	581.28 (.000)	-600.13 (.339)
Age[a]		295.04 (.022)**
Income 3[b]		-201.01 (.065)*
Education age[c]	-168.25 (.068)*	-289.59 (.007)***
Smoker[d]	260.81 (.074)*	317.42 (.032)**
H_motivation		251.18 (.049)**
Starting bid	.352 (.002)***	.313 (.007)***
Adjusted R^2	.138	.175

[a] 1>50,0<50. * p<.10.

[b] 1= 20-30K, 0 other. ** p<.05.

[C] 1<16, 0>=16 *** p<.01.

[d] 1 smoker, 0 other

As predicted, the higher the starting bid in the algorithm, the higher the mean mid-point WTP (*p*= .01). Respondents who left full time education on or after the age of 16 also had a positive effect on mean WTP. This result was significant for both tests.

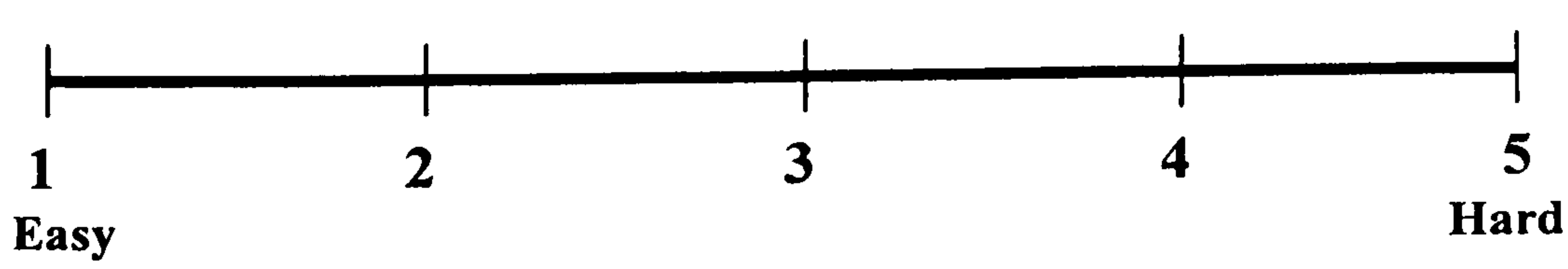
A higher WTP is elicited from the individuals that smoked in the study, for the FOB test, no other socio-demographic variables were significant. For the FS test, age and health motivation had a positive significant impact on the WTP (p = .05), implying that people who are older (> 50 years of age) and more health motivated are willing to pay more for this test.

An important result from conducting these two regressions is the level of R-squared associated with each model. The R-squared is very small telling us that the regression is only describing a very small proportion of the variation in the WTP data. This strongly suggests that there are other factors that the regression is not taking account of that must be influencing WTP.

9.5.8 Perceived difficulty of WTP task

After completing the WTP question the respondents were asked to indicate on a five point horizontal scale how easy or hard they had found the WTP question to answer (where one represents easy and five hard). The scale used for this question is presented in figure 9.3

Figure 9.3 Five point rating scale.



One-third of the sample circled the centre of the scale (3) claiming that they found the question 'OK' to understand and answer. Fifteen percent circled 'point 1' and said they had found the exercise straightforward and easy to deal with. Typical comments from the respondents that had found the task particularly difficult (circled ≥ 4.5 (9%)) were that 'I've never thought about treatments like these in terms of how much I am WTP for them' or 'once I had thought about it for a while I realised what you were getting at, but it was difficult to get my head round'. The overall response to this question is graphed in figure 9.4.

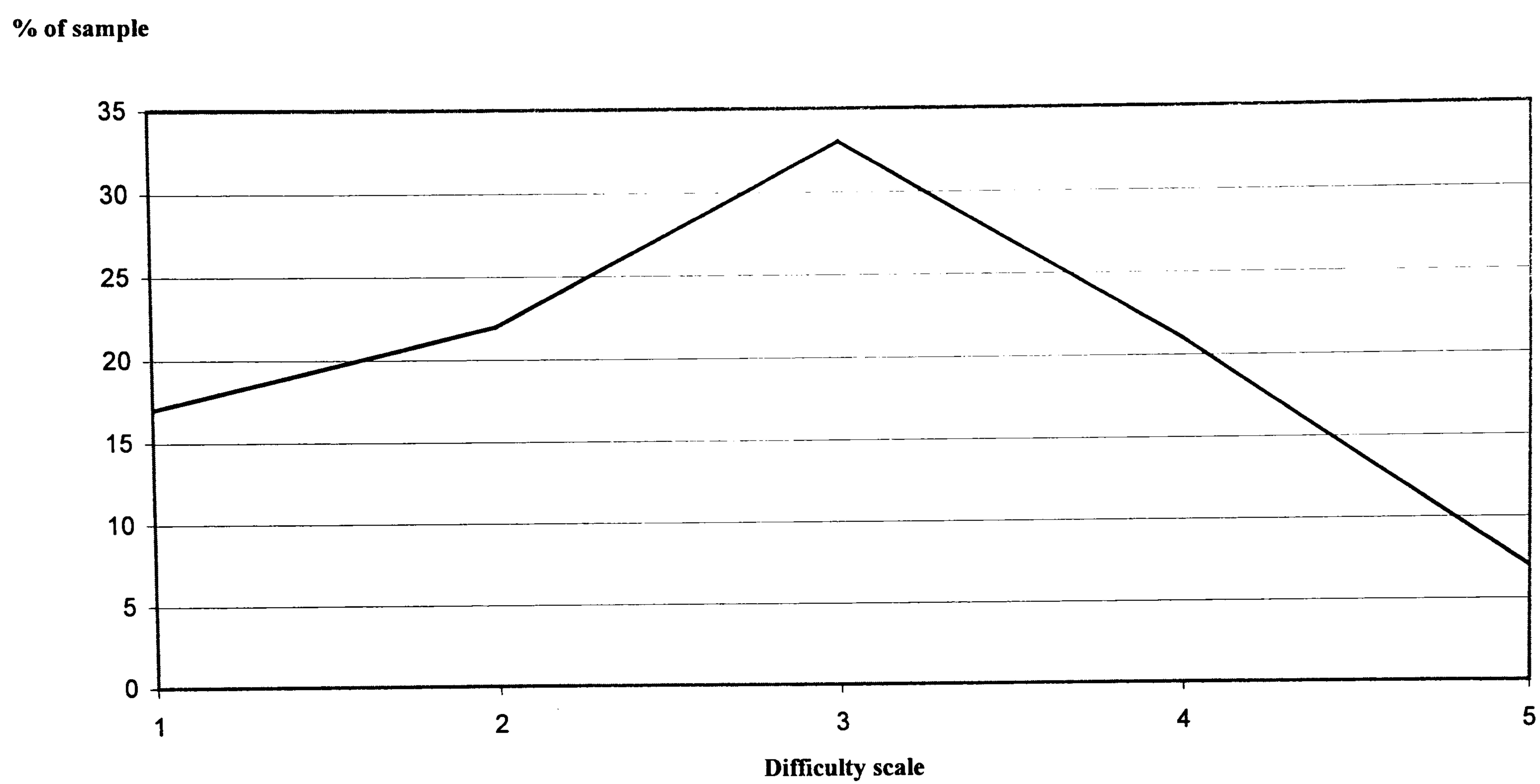
To establish if the ease of completion is related to the chosen algorithm the response to this question (by algorithm) is presented in table 9.9.

Table 9.9 Perceived difficulty of WTP question by algorithm

Rating scale	Algorithm 1 (%)	Algorithm 2 (%)	Algorithm 3 (%)	Total
1	6 (6.5)	8 (8.6)	2 (2.2)	16 (17.3)
2	6 (6.5)	7 (7.5)	7 (7.5)	20 (21.5)
3	13 (14)	9 (9.7)	10 (10.8)	35 (34.5)
4	4 (4.3)	6 (6.5)	9 (9.7)	19 (20.5)
5	1 (1.1)	1 (1.1)	4 (4.3)	6 (6.5)
total	30 (32.3)	31 (33.3)	32 (34.4)	93 (100)

Although the numbers are very small to make any broad conclusions about the effect of the algorithm on the ease of completion, it is clear that in category five more respondents in algorithm three found the task difficult to complete compared to algorithms one and two. Conversely, less individuals in algorithm three found the task easy (category 1) to complete compared to algorithms one and two. The majority of the individuals positioned themselves in the middle of the scale in all three algorithms.

Figure 9.4: Perceived difficulty of WTP question (1-easy, 5-hard)



9.5.9 Product elicitation

The next section of the interview aimed to encourage the subjects to consider the "opportunity cost" of the WTP amount that they had revealed for each test. To do this, the subjects were asked to think of a product/service familiar to them that was worth the WTP amount. Once that product/service had been established the subject was then asked if they would be willing to sacrifice it to have the test done, i.e. are you willing to give up that amount of money (or the value of that amount of money in terms of what else you could buy with it, i.e. opportunity cost). This question therefore puts the WTP question into a real-life context.

For the purposes of presentation, the products or services revealed have been divided up into categories; house, car, holiday, dinner/lunch, personal item. Examples of products put into each category are for house – furniture, new kitchen, windows; personal item – item of clothing, pampering session; the car category could also mean a car service or a MOT; the holiday item may relate to a two-week or a weekend break. So within each product category there were a wide variety of products offered relating to the value of the WTP amount. Table 9.10 summarises the results for each product category relating to the WTP value.

Table 9.10: Revealed products worth WTP amount

Agreed Amount		Product Category House	Car	Holiday	Dinner/ Lunch	Personal Item	Willing To sacrifice (%)	Not willing to sacrifice (%)
Agreed Amount	Bid	House	Car	Holiday	Dinner/ Lunch	Personal Item	Willing To sacrifice (%)	Not willing to sacrifice (%)
FOB TEST								
10			2		1		3 (100)	
50		1				4	5 (100)	
100		2	1			3	5 (83.3)	1 (16.7)
200		2	2	10		6	17 (85)	3 (15)
500		9	1	4		2	14(87.5)	2 (12.5)
800		3		1		1	5 (100)	
1000		12	2	10		5	26 (89.7)	3 (10.3)
>1000		2		2			4 (100)	
TOTAL		31	8	27	1	21	79 (89.8)	9 (10.2)
FS TEST								
10			1		1	1	3 (100)	
50		1	1			1	3 (100)	
100		1	1			3	5 (100)	
200		3	1	6		1	10 (90.9)	1 (9.1)
500		7		9		4	19 (95)	1 (5)
800		6		1			6 (85.7)	1 (14.3)
1000		11	2	10		4	25 (92.6)	2 (7.4)
>1000		1		2			3 (100)	
TOTAL		30	6	28	1	14	74 (93.7)	5 (6.3)

Across the two tests a total of 167 product/services were considered. The most popular items fell into either the house or the holiday category (FOB – 66% of items, FS – 73% if items). The majority of items were accepted as items of value that the respondent would be willing to sacrifice to receive a colorectal cancer screening test. This was an encouraging result in the sense that when subjects consider the opportunity cost of the WTP amount already stated, in the majority of cases, this reflected the value that they placed on the tests.

Fourteen products, worth the agreed WTP amount, were not willing to be sacrificed. Most of the time, this was because the respondent preferred the other test and were therefore not willing to give up their new piece of furniture, car service etc. for their least preferred test. This seems logical up to a point however what has to be borne in mind is that the respondents still claimed earlier in the interview that they were WTP that amount for their least preferred test. It became apparent after further explanation from the interviewer however that all but one respondent agreed that they would be willing to sacrifice the original products for the test (the one remaining respondent said that he would only sacrifice if he thought he had symptoms of the disease). Perhaps the unwillingness to sacrifice was more to do with a misunderstanding of the task rather than the elicitation of a WTP amount that was not accurate in terms of value.

Twenty respondents struggled with this task and could not think of a product worth the agreed WTP amount – this was down to a variety of reasons. There were three subjects when asked the direction of preference between the two tests that stated they would never have their least preferred test regardless of how much they needed it. However when it came to the WTP section of the interview, only two of them gave a WTP value of zero, the third respondent claimed a WTP value of £800 for the FS test (least preferred test). Yet when it came to the product elicitation section, this subject could not reveal a product and the explanation provided was that under no circumstances would 'he have anything shoved up his bum!' Another ten of the twenty respondents simply struggled with the WTP elicitation task and consequently found it difficult to reveal a product worth the WTP amounts.

Despite this group of individuals that struggled, it was still felt that this section of the interview proved invaluable in explaining how the WTP technique actually works. The respondents that were able to answer it found it really useful and it encouraged them to think of the opportunity cost of the value of the money.

9.6 Qualitative Information

This section of the chapter aims to use the qualitative information provided by the respondents to complement the WTP data. To begin, table 9.11 presents the mean WTP (mid-point data) values for the FOB and the FS test categorised by preference group.

Table 9.11: Mean WTP for preference groups

Preference Groups	FOB Test	FS Test
FOB	£662.27	£662.19
FS	£642.32	£706.55
No preference	£678.06	£674.17

To inform these WTP values it is essential to examine what was actually said during the interview by the respondents, in other words, given these WTP values how did they actually feel about these tests. To do this the analysis is broken down into parts. Taking each preference group in turn, the reasons for the WTP for FOB and then FS will be analysed separately. This will help to give some insight into why the respondents placed the value that they did on each test.

9.6.1 Preference for the FS test

Forty-five respondents told us that they preferred the FS test. As one would expect, this group produced a higher mean WTP for the FS test (£706.55) than for the FOB test (£642.32). Seventy-one percent of this group provided us with detailed reasons as to why they chose the FS test over the FOB test.

a. 'Quick Detection'

Nine people mentioned that they prefer the FS test due to the perceived 'quickness of detection'. These people liked the fact that the FS test would detect anything there and then and clear it up straightaway.

- *'I liked the once only aspect of FS',*
- *' You get the results straightaway. The FOB test would drag on for such a long period of time it would give me anxiety',*
- *'I prefer FS as I feel the doctors would be able to detect something straightaway and deal with it'.*

b. Convenience (time)

Ten individuals mentioned the word 'convenience' when describing why they prefer the FS test. They found the 'one-off' aspect of the test beneficial as they only have to endure the test once during their lifetime.

- *'I like the fact that the test is a one off, I would find that convenient as I am a busy man'.*
- *'The FS test is one off therefore it is more convenient'.*
- *'I prefer the FS test as I like the convenience of the technique'.*

Two people found the simplicity of the test attractive. The fact that they get it 'over and done with there and then' is an attractive feature of the test.

- *'The FS test is simplistic, the fact that I only have to get the test once during my life is worth more to me than the FOB test'.*
- *'I prefer the FS test as get it done over with quickly and I don't have to worry about sending stools through the post'.*

c. Sophistication

Sixteen people in total mentioned that they thought the FS test was a more advanced technique.

i. Professional present

Five interviewees commented that since there was a professional present during the FS examination they found that reassuring.

- *'Someone else is in charge, there is a reassurance that a professional is present'.*
  ' The FS test is a more sophisticated test as it is done by a professional… '.
- *'The FS test is the more professional of the two'.*

ii. Thorough test

Six people thought that the FS test was a more thorough test. This was because the scope examines the internal bowel.

- *'I prefer the FS test as it examines the insides of my bowels'.*
- *'With the FS test they can see inside you so there is more chance of spotting something'.*

- *'With the FOB test there is more risk of not detecting something'.*

iii. Advanced test

This category of comments referred to general statements relating to the perceived sophistication of the test. Three people fell into this category.

- *'Once you have had the FS test, you are clear for the rest of your life'.*
- *'The FS test is obviously the more advanced test'.*
- *'I feel the FS test is the better of the two'.*

d. Handling Stools

Finally, five people said that they preferred the FS test, as they do not like the idea of handling their own stools.

- *'Don't like the idea of stool samples, can cope with it once but not over 2 years'.*
  'I don't like the FOB test because I have to take stool samples, messy business'.
- *'It is a bit of a 'flaf' to do every 2 years'.*

9.6.2 Preference for the FOB test

Thirty-five respondents informed us that they preferred the FOB test. An interesting outcome for this group is that the average WTP estimated for both FOB and FS was equal (£662). Despite these equal WTP values, eighty-six percent of respondents from this preference group felt strongly enough to tell us why they had opted for the FOB over the FS test.

a. Embarrassment

i. Invasiveness of FS test

Nine respondents commented that the invasive nature of the FS test would put them off. One person said that,

- *'the invasive nature of the FS test would put people off and you have to think about that'*. Another said they would choose FOB because,
- *'it is less embarrassing than FS'*. Put quite simply these individuals felt that they
- *'prefer FOB as less embarrassing than FS'*.

ii. Laxative of FS test

Two individuals said that the laxative that is involved with the FS test would put them off having it,

- *'I'm not very keen on the laxative side of the FS test'*.

iii. Sex of consultant

The sex of the consultant performing the FS test seemed to emerge as an important factor when discussing aspects of the test. Three individuals mentioned that they would prefer a consultant of the same sex if they were given a choice and this would make it more likely for them to prefer the FS test.

- *'If I knew a man would be doing it then that would make me feel more at ease'* (male respondent),
- *'a female consultant would make me more willing to do the test'* (female respondent).

b. Timing of tests

Seven subjects chose the FOB programme, as they prefer the 'timing' of the tests. The starting age of the FOB screening programme was mentioned as an attractive feature as one individual said,

- *'I prefer FOB because of the starting age, I am worried about what can develop between 50 and 60'.* Others mentioned that they liked the fact they will be screened every 2 years,
- *'I prefer FOB because of the testing every 2 years over 24 years, I don't feel FS is thorough enough'.*
- *'FOB covers you for a longer period of time'.*
- *'I would rather do it every 2 years',*
- *'FOB goes on for a longer period of time therefore I prefer that one'.*

c. Travelling

Four individuals said that they would prefer FOB as that meant they did not have to attend their local hospital. One person commented that they were not mobile therefore did not find it easy to travel to hospital. Other comments related to general hospital features such as waiting time,

- *'the waiting time at the hospital clinic for the FS test would put me off therefore I would prefer FOB'.*

d. Home Test

Some subjects were attracted to the feature that the FOB was a home test. Seven people mentioned that they liked the privacy nature,

- *'I would prefer FOB as can do it at home and will not get flustered as can take my time,*
- *'I like the fact that the FOB test is done in the privacy of your own home'.* Another person simply said,
- *'I prefer a home test to a hospital test'.*

e. Pain and Discomfort

Nine expressed a concern about the perceived pain and discomfort associated with the FS test. Six of these subjects had received a sigmoidoscopy examination before and found it uncomfortable.

- *'I've had a colonoscopy before and I had a very bad experience with it, I found it very painful. I would need an anaesthetic if I were to have the FS test'.*
- *'I've had the FS test before and I found the after effects of it very painful as I was so constipated'.*
- *'I'm worried about the pain from having that test* [FS test]'.

One person said that they knew a relative who had the FS test before who had described their experience as being very painful.

'My sister has had that test before and she found it very painful. She is not the only person I know who has had a bad experience with that test [FS test], I'm worried that I will find it uncomfortable'.

One subject explained that they had had a 'bowel procedure' before (subject wasn't sure about the type of procedure it was) and as a result they found it very uncomfortable therefore would think twice about having the procedure again. They commented that they were *'left feeling very constipated'*.

Only one person mentioned that they would prefer the FOB test as they felt there is more scope for human error with the FS test.

9.6.3 No preference between the tests

Nineteen people had no preference between the two tests and eleven of them explained why. Comments made were largely similar, they were not put off by any of the aspects of either test nor did they have strong preferences for them.

- *'I recognise the benefits of both tests, I'm not really bothered to be honest'.*
- *'I'm not put off by any aspect of the two tests, I would do the first one that came in* [to the NHS]'.

An interesting outcome for this group is that even though they explicitly say during the interview that they are indifferent between the two tests, the WTP values reveal a slightly higher value for FOB over the FS test [£678 vs £674].

What is obvious from these results so far is that the WTP results are not coinciding with what respondents are actually saying about the tests. Only one preference group (respondents that prefer FS) are producing WTP values in the direction that one would expect. One possible argument is that the preference for the FOB test is present but is not strong enough to be picked up by the WTP values however this does not explain the 'no preference' group. Examining the comments that respondents made during the interview regarding the actual WTP technique might help to gain further insight.

9.6.4 WTP information from preference for FS group

Twenty-two percent of the group that preferred the FS test struggled with the WTP question. The main reason was the hypothetical nature of the technique and the idea that this procedure has the potential to save somebody's life.

Seven individuals tried to place a value on what their life was 'worth'. Even after explaining [again] the hypothetical nature of the technique the subject continued to believe that they were being asked to attach a value to their life.

- *'What is money worth to you when your lying in a grave. I would pay anything'.*
- *'That is such a difficult question and one that I can't really answer as this could save your life'.*
- *'How can I possibly do that* [answer WTP question] *– you can't attach money to peoples' lives'.*

Three individuals focused on how much they could afford to pay at the age in which they would be screened.

- *'I'm looking to my financial future at age 60 – this is how much I think I will be able to afford at that age'.*

Only one person in this group protested to the question,

- *'We have to think about society's WTP – you can't put extra charges on health care – we pay enough'.*

9.6.5 WTP information from the preference for FOB group

Eleven subjects (31%) that preferred the FOB test struggled with the WTP task. Two individuals claimed that they would be only WTP for the tests if they were particularly worried about the disease [misunderstood the concept of screening]. Unless they had symptoms of the disease they were not WTP. One person within this group protested to the style of questioning. Six misunderstood the aim of the WTP task and found it difficult to reveal a WTP value. Again, this was due to attempts to place a value on life, 'when it comes to life or death situations, you pay the money'. A further two people had affordability issues and claimed that pensioners are short of money therefore you have to make sure that everyone is able to pay.

9.6.6 WTP information from the no preference group

Nineteen individuals claimed they had no preference yet revealed a higher WTP value for FOB. Within this group, three struggled with the WTP question - two had affordability issues and there was one protestor.

The type of comments revealed by all the respondents that explicitly found the WTP technique difficult to answer are summarised in table 9.12.

Table 9.12: summary of WTP comments by preference group.

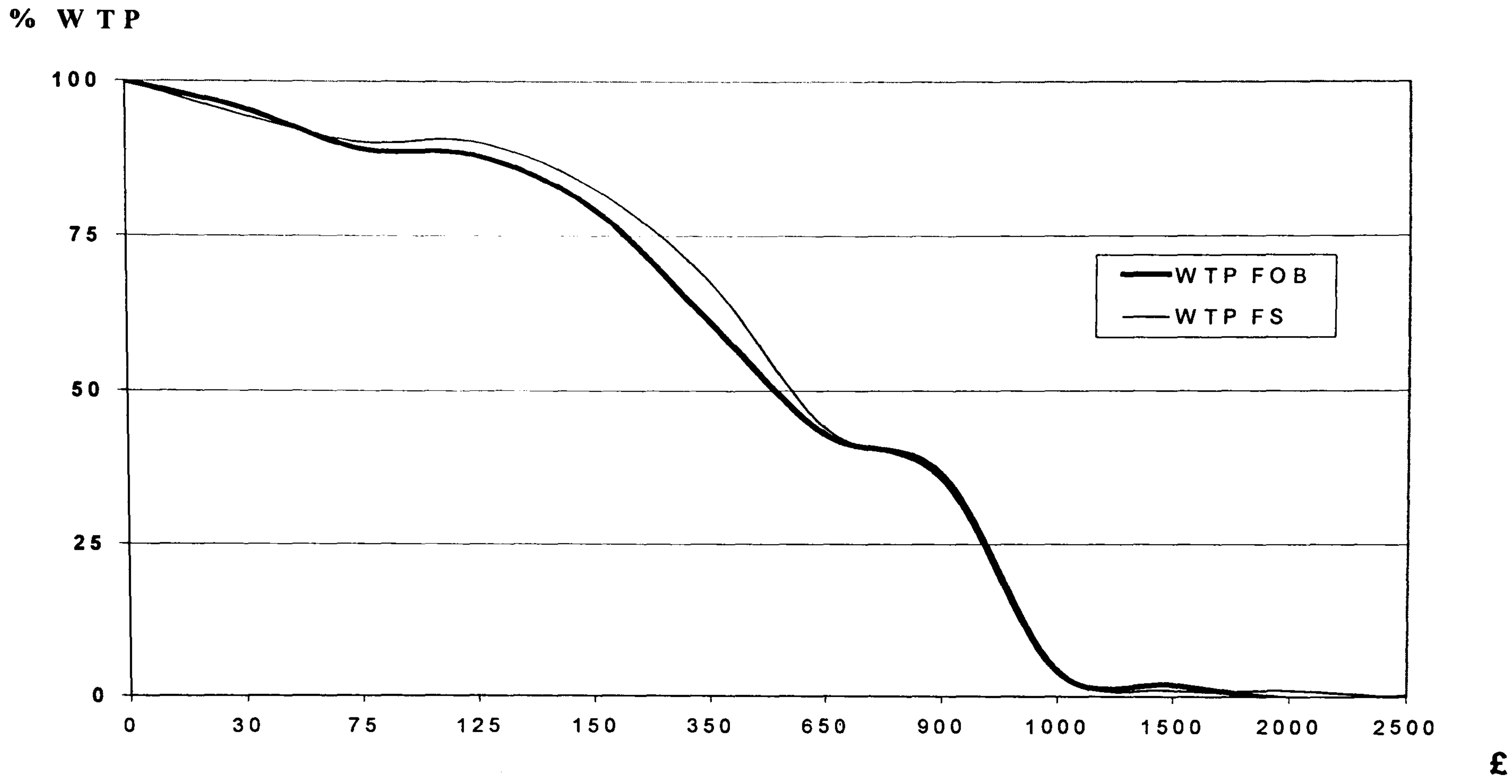
WTP Comments	Preference for FS test n = 45	Preference for FOB test n = 35	No preference n = 19
Number making WTP comment	16	11	6
Misunderstood WTP question	10 (22%)	6 (17%)	3 (16%)
Affordability issues	3 (7%)	2 (6%)	2 (10%)
Only WTP if thought had disease	0	2 (6%)	0
Protestors	1 (2%)	1 (3%)	1 (5%)

The number of respondents within each category are too small to make any broad conclusions concerning the WTP technique. It is clear however that the qualitative comments convey strong feelings felt by the respondents towards these tests and since the relative WTP values for the two tests were approximately equal, one may argue that the WTP values are not sensitive to these preferences.

9.7 Demand curves

Figure 9.5 illustrates the percentage of participants that are willing to pay (mid-point values) various amounts for the FOB and the FS test. As the graph shows the maximum WTP for the FS test is £2500 and £2000 for the FOB test. Most individuals' WTP for both tests were below £1000, on the other hand, 6% were WTP up to £2000.

Figure 9.5 : Demand Curves for FOB and FS screening.



9.8 Discussion

In the study presented in this chapter the WTP for FOB and FS is elicited using an iterative bidding approach conducted in an interview setting. A total sample of 106 individuals participated of which 100 said that they would be willing to have a test and 93 provided a WTP valuation for both tests. The FS test emerged as being the preferred test measured by both the 'direction of preference' question (46% preferred FS) and the WTP question [£687 (FS) and £659 (FOB)] – when the mid-point WTP values are analysed.

It is evident from the study results that there is a selection bias and possibly a hypothetical bias occurring given the overwhelming majority of the sample that indicated they would have a screening test. Recruiting members of the general population to be interviewed is always going to be a challenging task, it is inevitable that the more 'motivated, health conscious' individual will come forward. This kind of bias however is inherent in all kinds of research, i.e. taking the time to complete questionnaires, focus groups etc. The purpose of this study was to assess the 'process' of asking WTP and to measure the validity of the results given the rest of the information received during the interview. How well did the respondent understand and take part in the 'bidding process' and how valid is the WTP result given their reasons. These sorts of biases therefore are inherent within the previous OP/PS study and the CE study so comparisons in terms of the ease in which the WTP questions are answered is justified. Caution however has to be given to the nature in which these WTP results are going to be used for decision making.

The sample were divided into three groups, each received a different starting bid, £10, £200 and £1000. The results show that the bidding game approach is susceptible to starting point bias with the final bid amounts being influenced by the starting bids. Using extreme values as an example, for the FOB test, algorithm one produced a mid-point mean WTP of £418.67 compared to the £865.78 produced from algorithm three. Clearly the different starting bids of £10 and £1000 have influenced this result. The bidding game technique has been criticised as leading to starting point bias (O'Brien and Gafni 1996). However, as discussed in chapter four, health care studies that have applied the question format have produced competing results.

By using an interview to elicit WTP, the depth of information you get from each respondent is much greater compared to the data captured by mail questionnaires. This is the type of instrumentation technique that is recommended by Mitchell and Carson (Mitchell and Carson 1989), particularly when the study design adopts the bidding game format. The interview setting also facilitates the use of visual aids to communicate the good or product that is being valued and to explore any misunderstandings about what is being asked of the respondent. Unfortunately, however using an interview to elicit individuals' WTP is not without it's own drawbacks. An interview design can be costly and time consuming and the sheer fact that the interviewer is present throughout the elicitation process, may, in itself introduce biases. In this study, an interview was decided as the best means to bid with the respondents. The same interviewer conducted all 106 interviews therefore 'standardising' any biases that might have been present across all respondents. During the interview, respondents were encouraged to 'think aloud' enabling the interviewer to query any ambiguous comments that they made. The interview setting also allowed the researcher to encourage the respondent to think about the opportunity cost of their maximum WTP by getting them to think of personal items bought, or would like to buy, worth the agreed bid. This task proved fruitful in clarifying the WTP technique and getting the respondent to realise the personal value of the amount of money that they were claiming they were WTP. Out of 167 products revealed, there were 20 respondents who were not willing to sacrifice their chosen products. However after further explanation from the interviewer all but one agreed that they would sacrifice them for the screening tests. Getting the respondent to think of the opportunity cost of their WTP is always going to be a challenge in a hypothetical situation. However the product elicitation method did encourage the respondent to imagine actually spending the money on an item other than the screening service and therefore got them to think of the value of the money to themselves. The data collected from this section of the interview also helped to highlight the individuals who were struggling with the WTP task and even after further explanation still could not think about the opportunity cost of the money that they were claiming to be WTP.

The qualitative comments obtained through the interview helped to describe why respondents preferred one test to the other. Overall, seventy-three individuals explained their reasons for their preferences. The comments referred to many aspects of the tests from the timing, perceived sophistication, to a home versus a hospital test. It was obvious during the interview and from judging the depth of comments, that the respondents felt comfortable describing their feelings towards the screening tests. What is surprising, and perhaps a little worrying, is the insensitivity of the WTP question to pick-up on these preferences. Eighty-six percent of respondents that preferred the FOB test were happy to explain why, and their reasons seemed fairly logical. However this group still produced equal WTP amounts for both the FOB and the FS test. Perhaps the respondents were influenced by the presence of the interviewer (i.e. they felt under pressure to reveal a positive WTP amount for both tests). Maybe their demand for a screening test for colorectal cancer is so great that, despite their preferences for FOB, the FS test is still better than no test. This strategic behaviour has not gone unrecognised in CV studies. Mitchell and Carson do state that if the probability that a charge will be imposed for the service is low (as in the case of tax-financed screening service) then respondents are likely to overstate their WTP values (Mitchell and Carson 1989). Respondents realise that the health care good will be implemented if they give a high value but this will not affect how much they have to pay. If this is the case, then the participants in our study may have overstated their WTP values for both FOB and FS as they felt that this may have an influence of the likelihood of the service being introduced. However, as discussed in chapter four, there is little empirical evidence to support the hypothesis of strategic bias in CV studies (Bohm 1972; Smith 1979; Milon 1989).

Alternatively, the respondents may find it difficult to isolate one test from the other. During the WTP task, instead of considering the two alternative aspects of the screening test, the respondent may have considered screening of the disease generally and provided a WTP amount for that. Perhaps the respondents gained a feeling of 'warm-glow' from giving to a good cause and just provided a general WTP for the screening programme. This study highlights the importance of complementing WTP data with qualitative information that can be used to explain the direction of WTP values.

As in the case of the previous studies presented in chapters six through to eight, table 9.13 summarises the findings from this chapter.

Contrary to the results of the last two studies (open-ended and payment scale/ closed ended), the FS test is the preferred test in this study, but again, the demand curves for both tests are still very similar in shape (because the WTP values are very similar). Overall, the results suggest that subjects have a positive WTP value for colorectal cancer screening but it is not clear, from the WTP data, what screening protocol is preferred. A median WTP in the region of £650 does imply that subjects are WTP more than the resource costs for both tests. Put simply therefore, based on the results from this study, a colorectal cancer screening programme is demanded and the perceived benefits (judged by the WTP values) certainly would outweigh the cost but caution has to be given to the method used to elicit these results.

Table 9.13: Summary of findings from chapter nine.

Findings
1. 92% of the sample (total sample = 106) indicated that they would have a colorectal cancer screening test, if offered.
2. The majority of the sample opted for the FS test over the FOB test (46%).
3. Algorithms one and two produced WTP values for FOB and FS that were consistent with the overall preferences.
4. Algorithm three at first glance appeared to produce inconsistent WTP values (WTP FOB > WTP FS). However with further investigation it was discovered that the majority of this group actually preferred the FOB test therefore the WTP values were not inconsistent.
5. The results from the ANOVA test show that there is a statistical difference between the mean WTP values across the algorithms, both for FOB and for FS. This suggests the presence of starting point bias.
6. One-third of the sample circled the centre of the WTP difficulty scale indicating that they found the WTP task 'OK' to do.
7. The most popular items elicited that equated the elicited WTP values fell into either the house or the holiday category (FOB – 66% of items, FS – 73% of items).
8. All products elicited (eventually) by the individuals were willing to be sacrificed for the tests.
9. 20 respondents struggled with the product elicitation task and could not think of a product worth their elicited WTP amount.
10. 71 respondents provided detailed reasons as to why they preferred the FS test. In general, these reasons mentioned: - Quick detection of FS test. - Convenience (time). - Perceived sophistication. - Disadvantages of FOB test.
11. 86 respondents explained why they preferred the FOB test. These reasons covered: - The embarrassment of the FS test. - The timing of the FOB test.

- Convenience (home test).
- Lack of pain and discomfort.

12. 11 respondents explained why they had no preference between both tests as they recognised the benefits of both tests.
13. In general, a greater proportion of respondents that preferred the FOB test (compared to preference for FS test) struggled with the WTP task [31% vs 22%].
14. Reasons cited for finding the WTP task difficult referred to the difficulty of putting a value on life, the uncertainty of the future and one person protested to the WTP task.
15. The demand curves plotted from the WTP curves show that the demand for FOB and FS is very similar with the FS demand curve being slightly higher than the FOB demand curve in the middle range of the WTP values.

Chapter Ten - The Discussion

10.1 An overview

Routine screening for colorectal cancer is not available within the public sector in the UK however in the last 20 years there has been much research activity into looking at the feasibility of introducing two potential screening programmes: Faecal Occult Blood (FOB) and Flexible Sigmoidoscopy (FS) testing. Both screening protocols are described in detail in chapter five, briefly, the FOB test is a biennial home-test conducted between the ages of 50 and 74 years and the FS test is a once-only clinic-based test conducted around age 60 years. The majority of research to date focuses on the clinical and financial aspects of each screening protocol. There has been no research that has sought to elicit a value placed on each of these screening programmes from a general population perspective. To date, economic evaluations conducted have used 'cost per cancer detected' or 'cost per cancer prevented' as the outcome measurement. The contingent valuation method in health economics provides information that is more meaningful than these traditional measures when interested in the public perception of the screening tests. This is particularly useful in situations, such as this one, where the end outcome is very similar, i.e. early detection of colorectal cancer but the process of detecting that cancer is very different. The WTP values elicited for each test provides information on the relative preferences for these different processes of care.

The primary objective of the thesis is to contribute research evidence to aid the refinement of the contingent valuation methodology. In particular, emphasis is placed on the choice of elicitation method. A total of four elicitation formats are used in the thesis, all attempting to elicit WTP values for FOB and FS. The open-ended, payment scale (two scales), closed-ended and iterative bidding technique are each investigated in turn. Comparisons between each of the formats are made with respect to validity, reliability and bias issues that are all potential within a contingent valuation study and are discussed in chapter four.

To re-confirm the methodological components of the CV method that this thesis focuses on the reader is referred back to table 4.3 presented in chapter four (page 104). This table lists the main research areas that each chapter addresses. Chapter six looked at comparing the open-ended with the payment scale elicitation format and investigated the use of these formats with respect to consistency, criterion validity, protestors, ease of completion and the interpretation of zero responses. Chapter seven presents an extension of this original study by examining the effect of range bias within the payment scale format. The impact of this range bias is then assessed in relation to the original results presented in chapter six. Chapter eight looks at the closed-ended approach. Here, the affects of bid selection, consistency, ease of completion and the phenomenon of 'yea-saying' are examined. Finally, chapter nine presents the iterative bidding format. This study examines the process of asking WTP in a much greater depth as the WTP values are elicited by interview. Issues such as response rate, consistency, use of personal interviews, product elicitation and the 'warm glow' effect are discussed.

The study samples from all four studies are taken from the general population. Table 10.1 summarises the study protocol for each of the studies.

Table 10.1 Summary of study protocols.

	STUDIES			
	Open-ended/ Payment Scale	**Payment scale (new range)**	**Closed-ended**	**Iterative Bidding**
Target population	General	General	General	General
Data collection	Self-completed questionnaire	Self-completed questionnaire	Self-completed questionnaire	Interview
Administration	GP research network	GP research network	GP research network	Nottingham GP surgery
Environment	GP surgery	GP surgery	GP surgery	GP consultation room
Qualitative collection	'reasons' collected after both WTP questions for FOB and FS.	'reasons' collected after both WTP questions for FOB and FS.	'reasons' collected after each WTP question for FOB and FS.	In depth discussion about the WTP values after each WTP question.
Questions altered		None	None	One question inserted to encourage subjects to consider opportunity cost of WTP

The above table shows that there were minor variations in protocol between the studies, the main one being the change from self-completed questionnaires to interview questionnaires in the iterative bidding study. The reason for this was due to the nature of the iterative bidding question and the difficulty in administering that through a self-completed questionnaire. The collection of the qualitative data slightly altered between the studies as it was felt from judgement of the OP/PS dataset that asking subjects for reasons for FOB and FS separately would encourage thinking about each screening test in turn and therefore reduce the possibility of an embedding effect. Naturally, a more in-depth discussion about value and preferences was conducted in the interviews in the iterative bidding study. To take full advantage of the interview setting, one extra question was added to encourage subjects to consider the opportunity cost of their revealed WTP values. If the opportunity cost was considered too high, WTP for FOB and FS was consequently revised.

Nevertheless, despite these minor variations in study protocol the administration of the elicitation formats was kept exactly the same. Each respondent across all four studies received identical sets of information about colorectal cancer screening and screening in general. After the elicitation of the WTP values, the same questions were asked to collect information on socio-demographic details and attitudes towards health etc. This strengthens the thesis as it enables comparisons across all four formats to be made in terms of preferences and attitudes towards screening in general, the FOB and FS test in particular and the WTP values elicited.

10.2 The samples

Across the four studies, the sample populations range from a maximum of 2767 (on OP/PS study) to 106 subjects (in IB study). Table 10.2 presents a summary of the socio-demographic characteristics for all the respondents that answered at least one WTP question for all four studies. Since all of the studies are conducted in a primary care setting it is not surprising that there are a slightly higher proportion of females to males. The mean age is approximately 50 years and the majority of the study samples fall into the £10,000 - £20,000 household income bracket. With respect to colorectal cancer, about two-thirds of each study sample are worried about the disease although less than

Table 10.2 Characteristics of all study samples (that answered at least one WTP question)

Variables	Open ended/ Payment scale	Payment new range	Closed-Ended	Iterative-Bidding
n	2214	178	319	100
Gender:				
Female (%)	1720 (62.9)	104 (60.1)	209 (65.5)	51 (51)
Age (median yrs) [inter-quartile range]	49 [37-60]	50[39-59]	52[39-62]	55[49-65]
Income:				
<£10K	550 (24.8)	40 (27.6)	76 (27.6)	15 (17.6)
£10-20K	760 (34.3)	39 (26.9)	104 (37.8)	32 (37.6)
£21-30K	494 (22.3)	44 (30.3)	43 (15.6)	23 (27.1)
≥£30K	410 (18.5)	22 (15.2)	52 (18.9)	15 (17.6)
Proportion quite/very worried about bowel cancer	456 (20.6)	47 (26.5)	62 (19.4)	25 (25)
% who think chances of developing disease are higher than average	309 (11.7)	16 (9.1)	28 (9.1)	11 (11)
Number of visits to dentist in last 2 years (median) [IQ range]	3 [1-4]	4 [2-5]	4 [2-5]	5[2-5]
Age on leaving full time education (median)	16	16	16	15
Smokers (%)	19.6	46 (26.9)	66 (21)	12 (12)
Health motivation score (median) [inter-quartile range]	4.25 [4-4.5]	4.25 [3.5-4.5]	4.25 [3.75-4.5]	4.5[4-4.75]

15% think their chances of developing it are greater than average. Most respondents left full-time education at 16 years of age and visit their dentist, on average, 3-5 times every 2 years.

10.3 A traditional cost-benefit analysis (CBA)

Theoretically, the WTP method is applied for the use of measuring outcomes in monetary units in a CBA. Costs are then weighted against the perceived benefits to evaluate if the programme or intervention is worth implementing (subject to a budget constraint). To evaluate each of the screening tests in the context of a traditional CBA, the WTP values elicited in each of the four studies can be compared to the respective cost of running each programme.

It seems that, judging from the response to the first question asked in each study that an overwhelming majority of individuals demand colorectal cancer screening in general (88%-94%). In all but one study (IB study), the majority sample chose the series of FOB tests as their preferred option (in the IB study, the preferred option was the FS test). In retrospect, the only attributable factor that may have led to more respondents preferring the FS test in the iterative bidding study is the fact that the information and descriptions about the tests were delivered in an interview setting. Despite keeping the descriptions of both tests exactly the same, visual aids were used in the iterative bidding study and the interviewer's personal preference between these two tests is the FS test. It may be plausible that a bias towards the FS test came through in the descriptions provided to the respondents leading the respondents to consider that test more favourably than the FOB test.

If we use each of the WTP results at face-value for a traditional CBA, some interesting results arise. The cost of running each screening programme is estimated to be £62 per subject for the FOB programme [Frew, 2001 #13] and £56 per subject for the FS programme [Whynes, 2002 #14]. Comparing the WTP values from each of the four studies against these costs will lead to different policy decisions. If we take the WTP values revealed from the OP/PS study then the result is uncertain as the costs of each programme are very similar to the perceived benefit (WTP value). However if we look at the WTP values revealed by the IB and the CE elicitation format then the decision is much more clear-cut as the WTP values far exceed the estimated resource cost of each programme. Therefore in theory, with the IB and the CE study, the screening programme would be implemented with certainty.

Comparing costs with benefits seems, in principle, a straightforward exercise. However interpretation of these WTP values must be treated with caution. Chapter four discussed many aspects of the WTP methodology that are far from refined, aspects such as strategic bias, validity, elicitation format, even the formation of preferences are all contributing to the methodological debate surrounding the evaluation technique. This obviously raises questions about the validity of comparing WTP with costs as a tool for decision making. This thesis has been designed to examine some of the methodological components of the

WTP method by conducting four pieces of empirical work. The following section of this chapter summarises the overall findings.

10.4 Validity of the Study Design

The validity of a study design can be broken up into three parts; content, construct and criterion validity. Briefly, the content validity refers to the content of the survey instrument to ensure that the descriptions used within the instrument are unambiguous and meaningful to the respondents and do not influence the subjects response in either direction. The construct validity makes sure that the scenario being described to the respondent is correct within economic theory, for example, do the results predict the correct relationship between income and WTP. This type of validity can be tested by comparing the WTP values with values that have been measured using alternative outcome techniques such as the time-trade off or the standard gamble approach. Finally, the criterion validity refers to the verification process of testing whether stated WTP values are similar to revealed WTP values.

10.4.1 The content validity

The data capture instruments used in the studies were piloted and amended and ethical approval was obtained.

Before presenting the subjects with the questions, detailed descriptions about colorectal cancer screening and the FOB and FS programme were provided. The descriptions used deliberately excluded effectiveness information to equate to the descriptions that would be provided if the subjects were invited on a National Screening Programme. This meant that the results obtained were a good predictor of future screening compliance. A good area for future research could investigate the impact on the results from including effectiveness information in the descriptions.

The payment vehicle was clearly described within the questionnaire that under no circumstances would the subjects be expected to pay for either of the tests should one become available within the NHS. It was also emphasised clearly in 'bold type' that this was not a method for setting prices in health care.

The decision was made not to use taxation as the payment vehicle – taxation extends beyond valuing the good from an individual perspective and includes subjects' altruistic values towards society in general. Subjects are sensitive to taxation issues and asking them to provide a value using taxation may lead to 'protest' comments as subjects may feel that they are already paying through taxation for the NHS.

Asking subjects for a direct one-off payment means that concerns about the payment vehicle are, hopefully or at least more likely, to be brought up in the reasons rather than be inherent within the stated WTP value.

Each of the descriptions of the screening tests made it clear when in the subjects' lifetime they would receive each test. GPs excluded persons under the age of 25 years on the grounds of perceived irrelevance of screening for this disease in that age group.

10.4.2 The construct validity

The four studies produced results that are consistent with theoretical predictions. The variable representing the income of the subjects behaved as one would expect throughout. The income levels were categorised into four bands, <£10,000, £10,000-£20,000, £20,000-£30,000, >£30,000 and in all four studies, high income had a positive effect on WTP while the low income groups were more likely to provide smaller values.

Subjects that were particularly worried about the disease or thought that their chances of developing the disease were greater than average gave higher WTP values for each test in each study. It also seems as theory would predict that the more health motivated the individuals were, the greater the chance of higher WTP results.

10.4.3 The criterion validity

To check the criterion validity the study sample would have to be 'charged' for the screening tests to ensure that they are WTP what they say they are WTP. Obviously this cannot be done under an NHS setting.

Environmental economists often use estimates obtained from the travel cost method (revealed estimates) as a comparator for the stated estimates. In an earlier study conducted by the Nottingham research team, time and travel cost estimates were obtained for the FS screening test. Although these estimates were based on data obtained from a sample of respondents who had already visited the FS screening clinic (compliant individuals), it is interesting to compare the results to the WTP estimates produced in the thesis. Based on a sample of approximately 3,500 subjects drawn from twelve sites across the UK, median time and travel costs for a FS test were estimated at £19 per attendance, within an inter-quartile range of £11.8 to £29.4 [Frew, 1999 #6]. When compared to the stated values revealed in this project, these average revealed costs are of the same order of magnitude for the first study (OP/PS) and far below the stated valuations elicited by the CE and IB study. At face value therefore this shows that if we had only conducted the first study, using the open-ended and payment scale question format, we would have concluded that revealed WTP is approximately equal to stated WTP. However by doing the CE and IB study, we have shown that this result very much depends on the choice of elicitation format as in the CE and IB case the stated WTP is far greater than the revealed WTP.

10.5 The ordering effect

Theoretically, the response to the WTP question should not be influenced by the question order in a survey. In this thesis, the FOB WTP question was positioned before the FS WTP question in all of the questionnaire studies and in the interview study. Chronologically, the open-ended/payment scale (OP/PS) questionnaire was the first to be constructed and distributed, for consistency and comparability the order of questions were kept the same for the other studies. In the OP/PS study, the open-ended format produced a WTP value for the FOB test (£129) that was approximately £40 greater than the FS test (£86), however the median WTP values produced were exactly the same for both tests. The payment scale format produced means that were very similar for FOB (£93) and for FS (£91) and again the median WTP values produced the same value (£50 for both tests). The closed-ended study revealed similar WTP values for FOB and FS but in the opposite direction with the higher mean values elicited for FS (£1009) compared to FOB (£946). In the closed ended study the mean values are the median values. Finally, the iterative bidding study, depending on the algorithm, produced WTP values for FS that were similar to the FOB test, although the value was slightly higher for FS. The order of the WTP questions therefore does not seem to have a great impact on the magnitude of the WTP values.

The actual response to the WTP questions were marginally affected. The FOB WTP question produced a higher response rate compared to the FS WTP question for each of the elicitation designs. Table 10.3 summarises the findings:

Table 10.3 Response rate by elicitation design.

	Open-ended	Payment scale	Closed-ended	Iterative bidding
FOB WTP question (%)	70	86	92	100
FS WTP question (%)	65	77	86	100

The IB format produced a 100% response rate for both the FOB and the FS WTP question (due to the interview design).

It seems logical to assume that the higher response rate for the FOB WTP question is caused by the majority of respondents in all three studies actually preferring this test to the FS test. This makes it more likely that respondents will respond to the WTP FOB question compared to the WTP FS question.

According to the results therefore the order of the questions, although having a minimal affect on the response rate, does not affect the magnitude of the revealed WTP values. An area for possible future research could alter the order of the FOB and FS WTP questions to assess the impact on the results.

10.6 Value cues

10.6.1 Psychological anchoring effects

An anchoring effect occurs when the WTP value is influenced by pre-determined amounts within either the CE or the IB question design. Chapter nine presents the results from the IB study that clearly indicates that starting point bias is affecting the results. The average WTP values show a strong influence from the starting bid used in each of the three algorithms, £10, £200 and £1000. The difference between the mean WTP values is as much as £400 for the FOB test and £350 for the FS test. This outcome does raise questions about the validity of using the IB question given that the researcher can clearly affect the result by the chosen starting point within the algorithm. One suggested method of minimising this affect would design ten starting bid algorithms and assign 10% of the study sample to each bid algorithm (Eastaugh 2000). However this method would require a clear idea of the expected WTP distribution prior to conducting the study.

The 'yea-saying' effect potential within the CE design is also problematic. The final mean WTP value elicited from the closed-ended study was much greater than the mean WTP values elicited from both the open-ended and payment scale questions and even algorithm three (£1000 starting bid) in the iterative bidding question. Even with the aid of the WTP distribution obtained from the first OP/PS study to guide the bid selection, the highest bids (£100 and £200) within the CE design were not high enough to produce a substantial amount of subjects responding 'no' to these bids. Consequently the bids were revised and two higher bids added to the bid design (£500 and £1000). Even with these bids of £500 and £1000, approximately half the sample still responded with a 'yes'. This suggests that 'yea-saying' does indeed exist. It seems that no matter the value of the bid, a far higher proportion of subjects, compared to the other studies are responding with a 'yes'. Reasons that cause this yea-saying effect are not clear. Perhaps it is because respondents are asked to complete the questionnaire by their general practitioner whom they regard to be in a respectful position and the questionnaire has 'The University of Nottingham' written on the front which is an academic institution with an excellent reputation. This means that respondents are imagining that the bid values chosen are obviously amounts that they should be 'expected' to pay as they would not expect an academic institution to use bid values that are not realistic.

10.6.2 Range bias

When stated WTP values are influenced by the range of values chosen for the scale in a payment scale question design then 'range bias' has occurred. Chapter seven dealt with this issue directly by describing a study that compared two scales; a long scale incorporating values between £0 and £1000 and a short scale ranging from £0 to £100. Data for the long scale instrument was extracted from the original OP/PS study whilst new questionnaires were administered to collect data for the use of the short-scale instrument. Comparisons between the instruments produced marked differences in average WTP. The long-scale elicited a mean value of £90 and £92 for FOB and FS respectively, whilst the equivalent short-scale values were £60 and £58. Use of a longer scale therefore increases average WTP by as much as £30 for both screening tests.

This difference between the instruments is not as great however when median values are compared. The long scale instrument elicited a median WTP of £50 for both FOB and FS, the short scale instrument produced a median value of £40 (FOB) and £45 (FS). Therefore conclusions about the influence of the scale-range depend on the point of comparison; median or mean WTP values.

Whilst this is an interesting result within itself, it has important implications when consideration is given to the primary end-point of the original OP/PS study; comparison of elicitation formats. Conclusions were drawn from this study that the payment scale (when using the long scale) gives a higher median WTP compared to the open ended

(£30 (OP) v's £50 (PS)). If we had used the short scale range when comparing these two formats the difference would not have been as marked (£30 (OP) v's £40/£40 (PS)).

It is not entirely unexpected that WTP results will be influenced by the range of values chosen for the scale, however the essential point to note is that it can lead to different methodological conclusions. As discussed in chapter seven, 'standardising' a payment scale is not the solution as it would be either unnecessary as prior information concerning values is already known (hence no underlying need to research in first instance), or impossible as every health care intervention and population is different. It does raise concern however about the influence of the 'chosen' scale on final study conclusions.

These anchoring effects, discussed in section 10.6.1 and 10.6.2 must be factors to take into account when considering the choice of elicitation format as it is clear from the results presented in the thesis that they are having an influence.

10.7 Instrumentation technique

The self-completed questionnaires used in the OP/PS, PS (short scale) and CE study was an economical and practical method of administering the questions. This instrumentation method lended itself well to collaboration with the research network of GPs that distributed the questionnaires. Following a personal visit from myself, during which each practice was briefed on the required study protocol, a good response rate was achieved.

Time and effort had to be devoted to designing the questionnaires to ensure their suitability for self-completion. In particular, the following issues were addressed:

- Adequate background information provided at beginning to ensure that each subject answered the questions with same 'knowledge'.
- Information had to be kept to a limited nature to minimise cognitive overload.
- It had to be kept to an acceptable length.
- Had to ensure that the questionnaire is interesting to read to maintain attention.

The drawback of using self-completed questionnaires became apparent when analysing the data. Some subjects simply 'missed out' questions or clearly misunderstood what was being asked of them. It was also impossible to verify from the data if the subject had taken the time to read through the descriptions, therefore it was not clear if there was a full understanding of the screening programmes that they were being asked to value. Spending time revising descriptions to ensure understanding is meaningless if the subjects are going to fail to read it in the first place. There were also obvious cases within the data where a further explanation of the WTP method was required.

The iterative bidding study, largely to do with the nature of the question, used interviews to administer the questions. The same setting was used to conduct these interviews as the self-completed questionnaires in the previous studies (primary care setting). Clearly, the interviews provided greater flexibility in the communication of information (visual aids were used). There was a conscious effort however by the interviewer (myself) to use the wording and phrasing employed in the questionnaires. The interview did provide opportunities to explore preferences and values in more detail and also to explain any misunderstandings in the WTP methodology.

A disadvantage with the interviews however is that it may have been plausible that the responses provided by the individuals were influenced by the location of the interviews. It is possible that the subject may have become distracted by the fact that the interview took place in a GP consultation room with the interviewer being from the University of Nottingham. If the circumstances had been different, the response rate to the questions may not have been as high, also, the respondents perhaps would not have been as inclined to say that they would have a bowel cancer screening test nor give such positive values for both the tests.

The interviews were time-consuming to administer and as a consequence the sample size is smaller in the IB study compared to the other studies.

However the depth of information obtained from the interview study provided an opportunity to investigate more deeply the reasons behind the WTP values. Some of the discrepancies in the data within the questionnaire study were prevented in the interview study as subjects' views were explored in a much greater depth. Consequently, attitudes towards screening for colorectal cancer were revealed to a much greater degree in the IB study compared to the other studies.

The qualitative component of a WTP study is vital to the understanding of the results. It was clearly evident from conducting these four studies that the scope to collect this qualitative data is much greater when interviews are used. The quality of the information used to complement the WTP values in the interview study outweighed the large sample size and ease of data capture with the mail questionnaire studies.

10.8 Association of Ability To Pay (ATP) with WTP

The correlation between WTP and ATP can be taken as a measure of the internal validity of the methodology. The greater the ATP then the higher the expected WTP values.

Critics of the WTP method however are concerned about the influence of income on the monetary value placed on health care interventions. Kenkel's concern is that within a CBA the WTP responses depend on the distribution of income or wealth within the study [Kenkel, 1997 #15]. This means that if WTP responses are intrinsically affected by the distribution of income then this leads to evaluations 'intrinsically favouring programmes and diseases of the affluent over those of the poor' [Gold, 1996 #16].

Olsen et al. believe that if studies report an association between ATP and WTP then this is not, in itself, problematic. What is important however, when comparing WTP for two alternatives, is that the distribution of ATP is the same in both groups. For example, if rich people preferred good A and poor people preferred good B then we have the problem that the ATP for the goods is distorting results. However if there is the same proportion of rich and poor people in each preference group then the ATP does not become a problem (Olsen and Donaldson 1998).

This thesis has not been about the aggregation issue of WTP values. It is about the process of eliciting these values in the most valid and reliable manner ('the science of interrogation'). How to deal with the values once they have been elicited, although a very much related topic, is a separate matter. It is important however to check the influence of income on the elicitation of WTP values. For example, are respondents from a higher income bracket more likely to respond, less likely to provide protest answers etc. It is also a useful exercise to check that the ATP of the respondents is not distorting the WTP results.

In all four studies presented in this thesis, household income data was obtained. When comparing WTP between the FOB and FS programme, the effect of the ATP was checked to ensure that there was no distortion in results. Within each preference group, FOB, FS and no preference, the proportion of 'rich' and 'poor' were measured and chi-square tests conducted to verify that the differences between the 'rich' and 'poor' were not statistically significant. Table 10.4 summarises the results across all the studies:

Table 10.4 Influence of ATP by preference grouping for each elicitation format

Question formats *	Preference Groups n (%)			Chi-square	p-value
	FOB	FS	No preference		
Open-ended					
Rich	303 (60)	100 (62)	336 (67)		
Poor	204 (40)	62 (38)	168 (33)	5.298	0.071
Payment-scale (long scale)					
Rich	320 (58)	98 (61)	294 (64)		
Poor	228 (42)	63 (39)	165 (36)	3.361	0.186
Payment-scale (short scale)					
Rich	33 (43)	6 (33)	32 (48)		
Poor	43 (57)	12 (67)	35 (52)	1.225	0.542
Closed-ended					
Rich	84 (64)	29 (76)	77 (65)		
Poor	47 (36)	9 (24)	41 (35)	2.031	0.362
Iterative bidding					
Rich	18 (62)	16 (43)	12 (67)		
Poor	11 (38)	21 (57)	6 (33)	3.636	0.162

* 'rich' = £20-30K and >£30K; 'poor' = <£10K and £10-20K

The primary concern for the research question is that there are equal proportions of 'rich' and 'poor' between the preference groups and between the elicitation formats. Table 10.4 shows that the influence of ATP on WTP is small therefore ATP is not distorting the relative valuations between the FOB and FS programmes making it fair to make comparisons between the two tests and between the elicitation formats.

10.9 Sample representation

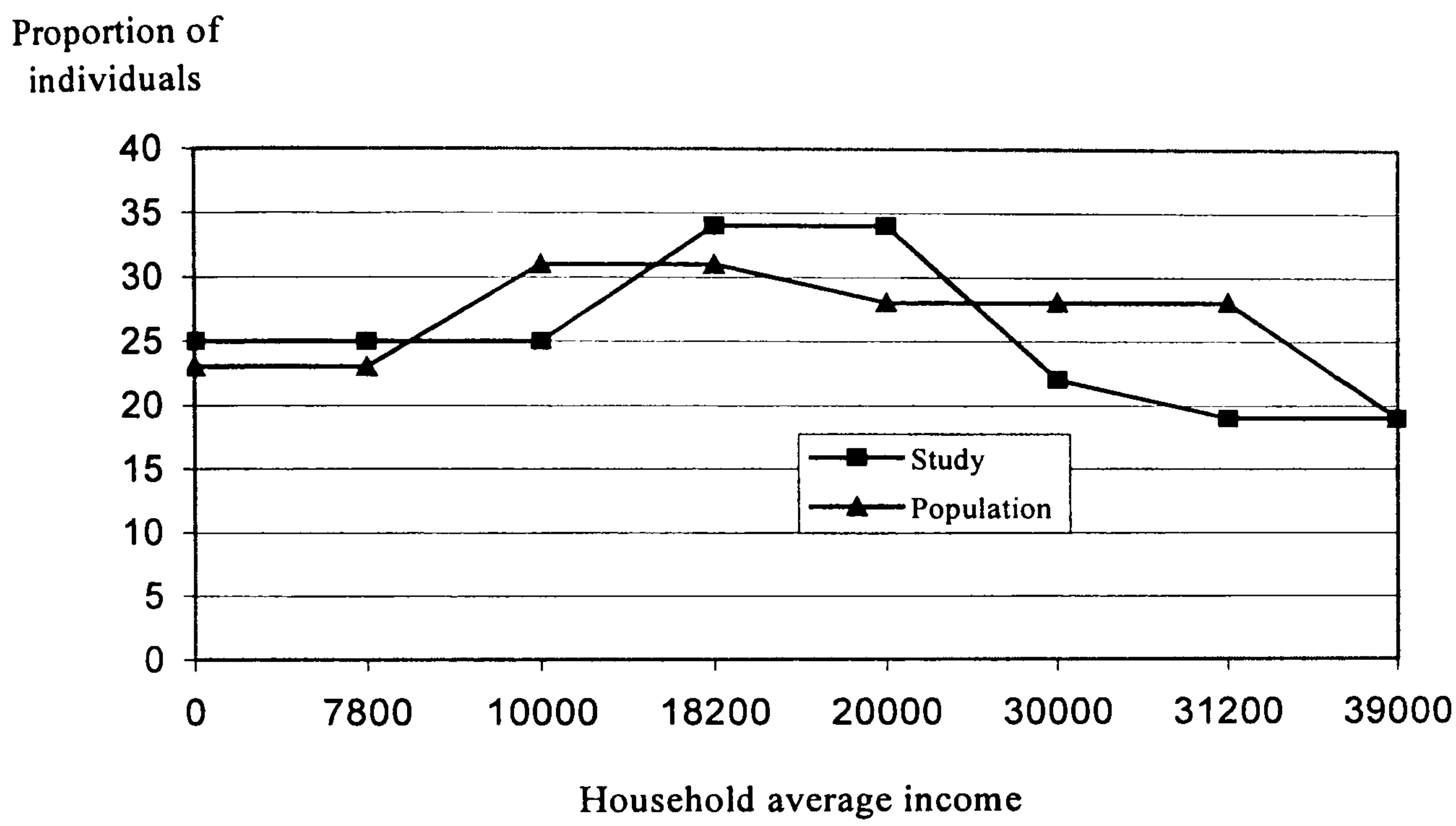
Theoretically, a CBA should take a societal perspective and incorporate all of the values from subjects who stand to lose or gain from the health care programme being valued. This will include all the subjects who benefit indirectly as well as directly so as to ensure the measurement of any externalities produced. Obviously measuring values from all subjects is impractical so a representative sample is usually targeted.

Each colorectal cancer screening programme is designed to target a different age group, 50-74 years with FOB testing and approximately 60 years with FS testing. WTP for these programmes is estimated from a sample taken from the general population therefore individuals from a much broader age band are targeted. This results in the value of the service being measured from the broadest perspective as it takes into account 'potential' users of the service and incorporates any altruistic values from the service existing. The data for all studies was collected from samples gained from a primary care setting within the NHS, in total 3232 questionnaires and 106 interviews were completed.

In general, research at the primary care level presents two challenges for sample representation. First, the Fourth National Morbidity Study states that social classes IV and V between the ages of 16 and 64 years have an increased likelihood of consulting their GP [OPCS, 1991-1992 #11] and second, it is also common to see females visit GPs more frequently than males.

As reported throughout the thesis, the proportion of females sampled were slightly greater than the males. When the study samples are compared with the East Midlands population census statistics there is an accurate representation of income groups. Figure 10.1 displays the relationship between the distribution of household income from the first study (open-ended/payment scale) along with the Trent population income data [ONS, 1995-1998 #12] and illustrates that the two income distributions lie close to one another. This is an important outcome given the controversy surrounding the potential influence of ability to pay on willingness to pay discussed in the previous section. Table 10.2 shows that there is no real difference between the socio-economic characteristics and in particular the income groups, across the four studies therefore comparisons of the WTP values elicited by the different question formats is justified.

Figure 10.1 Income distribution for study sample (open-ended/payment scale)



10.10 Embedding effect

An embedding effect influences results when individuals consider the value of the global good rather than considering the individual aspects of the alternatives. In this case, individuals would provide a WTP for colorectal cancer screening without considering the different protocols of the alternative screening programmes.

To begin the study, respondents were asked for their direction of preference between FOB and FS with a no preference option. Overall, the results predicted that approximately 40% of the sample preferred FOB, 13% FS and 40% had no preference (the exception was the IB study with 20% opting for FOB, 40% FS and 30% had no preference).

The message given from the WTP values do not reflect these preferences. If we look at the median values taken from the OP/PS study then estimated WTP for both FOB and FS is £30 for the open-ended and £50 for the payment scale. It is fairly safe therefore to presume that the only aspect of the valuation process that is influencing these values is the design of the elicitation format. The other studies produced similar results. Median WTP from the payment scale (short-scale) was estimated to be £40 and £45 for FOB and FS, for the CE study the values were £946 (FOB) and £1006 (FS). Finally, the IB study estimated WTP to be £646 for FOB and £724 for FS. If we had no knowledge of the direction of preference between the tests and there was only the WTP value to guide us then we would conclude that there was no real strong preference for FOB over FS (or vice versa). It is only with the preference group information and more importantly the reasons behind these WTP values that evidence of strong preferences are given.

Closely related to the 'embedding effect' theory is the 'warm-glow' hypothesis. It is possible that the respondents feel some moral obligation towards the screening programme going ahead and therefore provide a WTP value that makes them feel they are doing their 'fair share'. This is a public health care programme that is being valued therefore individuals may have felt propelled to do the right thing and provide a positive value irrespective of the screening test. One possible area for future research would examine the impact from asking respondents firstly, for a total WTP for colorectal cancer screening in general followed by a WTP for FOB and for FS screening individually. This would encourage respondents to divide their <u>total</u> WTP among the two screening programmes.

Possible explanations therefore for the apparent similarity between the relative WTP values for FOB and FS is that a possible 'embedding effect' occurred and/or subjects were influenced by the 'warm glow' feeling.

10.11 Strategic bias

The screening programmes for colorectal cancer are being evaluated in the context of a tax-financed health care system. There is potential therefore for the subjects to respond strategically as they realise that they will not actually have to pay for the service but it is more likely to be implemented if they provide a high WTP value. Alternatively, the subjects may misunderstand the task and become 'offended' by the question providing a protest response. With a protest response an unusually high or low value or a zero value is offered. This is because the respondents object to being asked to 'pay for health care twice'.

Interpreting WTP values is difficult without reasons to provide an explanation. The reasons are used to determine if the response is a 'protest'. In the first open-ended/payment scale study, 19% of the sample gave a 'protest' reason for their WTP value. These reasons were broadly similar with comments such as 'the NHS should bear the cost' and 'having paid taxes, one shouldn't have to pay more'. The proportion of the study samples providing 'protest reasons' were similar across all the studies in the thesis (19% for the payment scale (short-scale), 10% for FOB and 14% for FS in the closed ended study). The iterative bidding study produced a lot less 'protest' comments mainly due to the fact that the study took place in an interview setting with the interviewer clarifying any misunderstandings about the WTP method.

The question rests therefore on how one should deal with these protest comments. Advice provided by the literature is contradictory. Diamond and Hausman (1994) argue that it is standard practice to eliminate all 'protest zeros' as this type of zero is not a credible answer [Diamond, 1994 #18]. Dalmau-Matarrodona (2001) however argue that it may not be correct, statistically, to remove all respondents who have stated a protest zero as removal of these respondents will incur a potential cost to the loss of information [Dalmau-Matarrodona, 2001 #19].

Chapter six presents what happens to the overall study results when 'protestors' are removed from the analysis. The results are interesting in that the removal of this group of respondents made very little difference to the overall median WTP for the open-ended question and absolutely no difference to the payment scale results. If by removing the 'protestors' the results of the study do not alter then this casts doubt as to whether excluding respondents that protest is really required.

Removing protestors from the analysis is in my opinion, methodologically wrong as it weights any protest at zero. It is much better to leave the protestors in the analysis and assess the affect the group is having on the overall results (as done in chapter six). The protestors are only identified by the comments made in the qualitative section of the study therefore how does the analyst know how many other respondents have protested without informing them? What might be even worse is if respondents who have protested without making any comment put a 'lesser' WTP value than their 'true' value and these 'lesser' values are analysed as 'normal, valid' responses. We have no way of identifying this group and distinguishing them from the rest of the sample. Removing the 'protestors' (identified by comments written/spoken) therefore does not necessarily eliminate <u>all</u> protestors from the sample.

Another interesting point to note is that protesting respondents are often referred to as individuals that have misunderstood the objective of the task. What might be the case though is that these respondents have actually thought about the task in greater depth compared to the rest of the respondents. Consequently, they have taken the task more seriously and because of the nature of the question and the particular thought process of the individual the subject may answer strategically. It is plausible that respondents who do not protest have answered the question without giving it much thought and have put down 'any old value' just to get through the questionnaire/interview. Protestors have obviously taken the time to write comments to support their protests. Although it is very difficult to identify this sort of response when analysing results it is a plausible point and one that should be noted when the advice in the literature seems to be that all protestors should be excluded.

The real focus therefore should be on trying to communicate the nature of the exercise more effectively. Trying to ensure that the information contained within the questionnaires is explained fully is challenging enough but there are no means of actually verifying that respondents have fully read and understood the information. For WTP exercises to be facilitated in the most efficient manner therefore they have to be administered in an interview setting to ensure that the respondents are really understanding the task presented to them. With properly trained interviewers the true 'protestors' can be identified during the process of the interview rather than identified (or perhaps even 'missed') by the analyst from qualitative comments provided in a questionnaire.

10.12 Preferences in relation to WTP

The WTP technique measures how much individuals collectively value a good. It is an ideal method to adopt when valuing two alternatives that achieve the same clinical outcome but with different processes of care, i.e. the FOB versus the FS test. There are four studies presented in this thesis that measure the value placed on FOB and FS screening.

The primary aim of the thesis is to investigate the methodological qualities of the WTP method by comparing the elicitation formats. Although it is important to keep this in mind, in order to do this efficiently, a wider consideration of the preferences revealed in the studies have to be considered. In other words, does the direction and magnitude of the WTP values match what the individuals are saying in other parts of the studies?

Table 10.5 summarises the mean and median WTP values for both the screening programmes.

Table 10.5 Willingness To Pay for FOB versus FS screening.

WTP		Open-ended	Payment scale Long-scale	Payment scale Short-scale	Closed-ended	Iterative bidding
FOB	Mean	£129	£93	£60	£946	£659
	Median	£30	£50	£40	£946	£650
FS	Mean	£86	£91	£58	£1009	£687
	Median	£30	£50	£45	£1009	£650

Preferences for FOB and FS can be derived from looking at the overall level of WTP for each test across the studies. As mentioned previously, all but one study produced an order of preference that was greater for the FOB than the FS programme (the iterative bidding study was the exception). Logic would therefore predict that the WTP for FOB, in these three studies, would be greater than the WTP for FS.

Again, it is the different messages conveyed by the mean and median WTP values that produce an interesting outcome. In the open-ended, payment-scale (long and short scale) and in the closed-ended study the mean WTP value for FOB is greater than FS. The greatest difference is elicited in the open-ended study with a WTP value £159 for FOB versus £86 for FS. The relative median WTP values convey a different message. In the open-ended and payment-scale (long scale), the median WTP values for FOB and FS are exactly the same whereas in the payment-scale (short scale) the direction of magnitude has changed with the median WTP for FS (£45) being greater than FOB (£40).

The iterative bidding study produced mean WTP values that were higher for FS compared to FOB in the manner that one would have predicted (given the original preference direction). The median WTPs however are exactly the same for both tests.

This raises the question, yet again, of whether one should use the mean or the median values? It seems surprising that despite the strong preferences revealed by the direction of preference question, the median WTPs elicited for each test is very similar, if not exactly the same in all four studies. It is only when the mean WTP values are analysed do differences occur. It is common and certainly the case in these studies, that WTP distributions are highly skewed with the majority of the sample eliciting values that lie to the left of the distribution. The mean values are therefore influenced by the small numbers of very high WTP amounts. Although the mean values take these high amounts into consideration, overall this may give a false impression of what the majority of the public are WTP. The median values however are not influenced by the high WTP amounts but obviously fail to encompass the high values that may be genuine indicators of strong preferences from a minority in the study sample.

From looking at the WTP figures, it is clear that there is a preference for FOB over FS for all but one study (iterative bidding study). Establishing the reasons for these preferences is paramount to understand and inform the WTP structure. The explanations provided by the respondents about either of the two tests and about the WTP method itself helped to gain an overview of the public perception of the screening tests and the valuation technique. The following table provides an overall summary of all the comments made over the four studies.

Table 10.6 Summary of comments

	Preferences for:	Comments
1	Overall screening	• Think screening tests are vital.
		• Peace of mind
		• Screening tests incur cost-savings to the NHS.
2	FOB programme	• Find the FS test too invasive.
		• Don't like the idea of a laxative with the FS test.
		• Like the privacy of the home-test.
		• Like the frequency of testing every 2 years for peace of mind.
		• Concerned that an 'abnormality' may occur before age 60.
3	FS programme	• The FS test seems more sophisticated than the FOB test.
		• Get reassurance from professional being present during procedure.
		• Get it over and done with at age 60 years
		• No need to bother with home–test, sounds too much hassle.

The above reasons help to complement the WTP values and explain to the policy decision maker the specific reasons why individuals are WTP what they are for the tests.

The reasons also helped to predict the direction of WTP in the regression analyses done in the studies. Chapter six presents two sets of regression results, one to predict the likelihood of responding, and the second one to explain the distribution of WTP. It seems that if the individuals expressed difficulties in being WTP, found the WTP task not applicable to them or protested, then they were less likely to provide a WTP response. If they felt that colorectal cancer screening was beneficial and had family experience of the disease then they were WTP more for both FOB and FS.

The iterative bidding study went one-step further and as well as getting explanations for response, 'tested' the individual to see if they were willing to sacrifice what they had said they were WTP. Products revealed ranged from items for house such as furniture etc, a car or related product, e.g. car service or MOT, a holiday (weekend break or 2 weeks), dinner or a personal item such as clothes. In total, there were 14 products worth the stated WTP amounts that were not willing to be sacrificed and as a consequence the WTP had to be revised.

10.13 Consistency of results with NOAA guidelines

There are no formal guidelines on how to conduct WTP surveys within a health care context. In the field of environmental economics though, the National Oceanic Atmospheric Association (NOAA) published guidelines providing recommendations on how to measure values of environmental goods. These guidelines are presented in detail in chapter four however it is interesting to see how the results from the studies in this thesis compare.

The main focus of the thesis is the comparison of how well the alternative elicitation formats perform when eliciting WTP values. NOAA recommends the use of the closed-ended format as it is believed that the design resembles decisions that individuals have to make in real-life. NOAA dismisses the use of the open-ended design as it is thought to lead to over-statement of values, and the payment scale design because of the possible anchoring effects.

In this thesis, chapter six presents results that show that the payment scale design elicits slightly higher median and very similar mean WTP values compared to the open-ended design. Chapter seven directly investigates the possibility of an anchoring effect (range bias) within the payment scale design, and results suggest that the WTP values are indeed being influenced by the scale. Whilst the results therefore agree that the payment scale WTP values are influenced by the range, the open-ended design certainly did not produce vastly greater WTP values.

The closed-ended study, presented in chapter eight, has its own set of biases. The results suggest that 'yea-saying' is occurring with significant proportions of subjects saying 'yes' at the higher bids. Overall, when the closed-ended mean WTP values are compared to the open-ended and payment scale WTP values, the magnitude of difference becomes apparent. The mean values elicited by the closed-ended question were up to seven times higher for the FOB test and eleven times higher for the FS test.

The iterative bidding format presented in chapter nine, suffered from starting point bias as the WTP values were strongly influenced by the bid algorithm. Overall, the WTP values are greater than the open-ended and payment scale values but not as high as the closed ended design.

It seems therefore that each elicitation format has its own advantages and limitations. The closed ended is prone to 'yea-saying', the payment scale is affected by range bias, and the results from the iterative bidding question are determined by the algorithm. The open-ended is not prone to any specific bias but due to the lack of guidance within the question design, it is difficult to ascertain if the WTP amounts are reflecting true values.

It is important to mention that when NOAA recommended the closed-ended format in an environmental context, they did so by recommending it to be administered in an interview setting. Future research would have to compare the values from the study presented in chapter eight, where a questionnaire was used, to the values elicited in an interview to see if this makes any difference to overall conclusions.

The main lesson to be learnt from conducting these studies for the thesis is that WTP surveys should be conducted using interviews. This is what NOAA recommends in the guidelines. The main advantage being the opportunity to collect qualitative data to complement the WTP values. Qualitative information aids the researcher to draw conclusions about the study and to explain reasons why respondents are WTP the amounts they have provided.

10.14 Overall findings

The objective of the thesis was to contribute to the development of WTP as a tool for measuring values from the general population. The main aim was to examine the effect of using different elicitation formats, the following question designs have been investigated:

- Open-ended
- Payment scale
- Closed-ended
- Iterative bidding

The project involved asking samples drawn from the general population what their WTP would be for two potential screening programmes for colorectal cancer; biennial FOB testing and once-only FS testing.

Overall, the main findings of the thesis have been:

1. That the hypothetical nature of the WTP methodology necessitates the need to communicate information and administer questions in an interview situation.

2. The importance of using the qualitative information to explain WTP values is such that this data needs to be collected in an interview with maximum opportunity to explore preferences and values.

3. Using a payment scale question design leads to a higher response rate and higher overall median WTP values compared to an open-ended question design.

4. The payment scale WTP values are greatly influenced by the chosen range of values in the question design with a smaller range producing lower average WTP compared to a longer range.

5. The open-ended and payment scale produce broadly similar valuations whereas the closed-ended question design produced significantly higher WTP values and different justifications for these valuations.

6. It is suggested that 'yea-saying' effects are influencing the closed-ended valuations.

7. Use of the iterative bidding design produces significantly higher WTP valuations compared to the open-ended and payment scale formats but not as high as the closed-ended format.

8. Using different initial bids to generate the bidding process in the iterative bidding format suggests evidence of starting point bias in the bidding game.

9. Encouraging respondents to think of the opportunity cost of their WTP valuations helps to put the WTP task into context.

10. There is a slight evidence of an ordering effect in terms of the response rate to the WTP FOB question and the WTP FS question, however no evidence of an order effect in terms of the magnitude of WTP revealed for both tests.

11. There is no influence of ATP on the relative WTP valuations for FOB and FS as there are equal proportions of 'rich' and 'poor' in each category.

12. Where median WTP values are analysed, all four elicitation designs produced very similar valuations for FOB and FS suggesting a possible embedding effect.

13. The median WTP values failed to capture strong differences in direction of preferences.

14. The mean WTP values reflected the difference in direction of preference between FOB and FS but were greatly influenced by small numbers of high valuations.

Now that all the empirical work has been completed for the thesis, it is useful to pull the research findings together for the four studies and use them to recommend how I would design a WTP study now. Therefore based on my experience, from doing the work for the thesis, I would recommend the following study design if tasked with the objective of estimating the value of two alternative tests using a WTP method:

- Fundamental to the study, I would recommend the use of interviews as the vehicle to collect the data. Use of visual aids to ensure an understanding of the goods being valued and a thorough explanation of the WTP technique to minimise protest responses. Respondents may find it easier if examples of previous WTP studies are used to describe the WTP method, this may help to minimise any suspicious thoughts that they are having about the study attempting to set 'charges' for health care. If the individuals understand how the WTP values are used, once they are elicited, then they may understand and accept the task a lot easier.

- Before moving on to the WTP section, I would ask for an ordinal direction of preference by getting the subjects to choose their preferred alternative. This question should always include a 'no preference' option. This is important as only using the WTP values to measure preferences may lead to wrong conclusions (as shown by the results of this thesis where a similar WTP value had been provided for each test despite one test being preferred to the other and strong qualitative comments provided indicating direction of preference).

- To gain an understanding by how much the individual values the global good, i.e. WTP for the overall treatment as opposed to the individual alternatives, I would start by using an open-ended question to ask total WTP. Reasons for the open-ended question design are because it is the question format that has the least amount of 'props' or guides and would start off the process of the respondents thinking by how much they value in the context of WTP. It may be seen as a difficult cognitive task to ask individuals to do, but in the context of an interview setting the interviewer is there to clear up any misunderstandings.

- Following the open-ended design, I would use a payment scale question to ask WTP for each alternative. The payment scale question design, although prone to range bias, does not require a prior understanding of the expected WTP distribution (as in the closed-ended design) nor does it require the construction of a complex algorithm structure to minimise starting point bias (as in the iterative bidding format). As long as the range is 'realistic' to the good being valued then the payment scale question design is by far the easiest method to administer. In theory, because the respondent has already elicited a total WTP value for the global good, the payment scale question should then help the respondent to divide that total WTP between the two goods therefore minimise an embedding effect.

- When designing the interview format the study should 'swap' around the order of the WTP question between the two alternatives to assess the impact of the ordering effect.

- Once the WTP values have been elicited there needs to be adequate time devoted to a structured qualitative section that asks individuals for their reasons for their WTP. This will help 'validate' the WTP amounts and provide a clearer understanding behind the value placed on the good.

- To check that the individual is actually WTP the amount that is stated in terms of the potential opportunity cost of that money then they should be asked to think of a product or a service that is worth the amount they have stated they are WTP. They should then be asked if they would be willing to sacrifice that good for the alternative being valued. If they are reluctant to do this then the WTP should be revised.

- A series of socio-economic and demographic questions needs then to be asked to gain information about the subjects personal profile.

Although the studies presented in this thesis cover a lot of issues and produce some interesting findings, there are still some key areas that require future research to shed further light on the 'appropriate' elicitation format.

At the time the original questionnaire was constructed for the OP/PS study, the effectiveness information for the Flexi-scope test was not available. There was also a concern about how well the subjects were going to understand all the information presented to them about colorectal cancer screening and the WTP method. The questionnaires therefore did not contain information concerning the differences in effectiveness between the two screening programmes. It would be interesting to see how the magnitude of the relative WTP values would change from the introduction of this information so future work would re-administer the same questionnaires with this information added to test the impact on the results.

The response to the WTP for FOB was slightly greater than the response to the FS question. Although the difference in response was not huge it would still be interesting to vary the order of these questions and assess the impact on the results.

The WTP amounts elicited for FOB and FS were very similar and in some cases the median WTP values were exactly the same. New questionnaires could be designed with a total WTP question before the individual WTP questions for FOB and FS. A total WTP for colorectal cancer screening could then be estimated followed by the individual WTP values for FOB and FS.

The closed-ended question design produced values that were much higher than the alternative elicitation formats. The question however was not administered using the recommended format (questionnaires instead of the recommended interviews). Future work could ask the closed-ended question in an interview and then assess the impact of using a different instrumentation technique on the results.

10.15 Research within the CV field

The level of research interest into the CV technique within health care has risen dramatically over the years. Consequently, it has become a dynamic area of research within health economics with new recommendations, reviews and suggestions arising frequently in the literature. One such movement has been the suggestion of the marginal WTP elicitation format [Donaldson, 1998 #20]. Using the marginal approach, individuals are first asked what treatment or service they prefer and then asked what their maximum WTP would be to have their preferred option over their less preferred option. Therefore instead of the absolute WTP being important, it is the relative WTP that is interesting. In the studies presented in this thesis, it is the absolute WTP that was being measured and then compared between the two screening programmes. It is the difference in the absolute WTP that was the interesting outcome. At the time the original OP/PS study was being designed, the marginal WTP approach was only starting to be advocated in the literature by Donaldson which is why it was not incorporated into the original research design. To enable direct comparisons with the original OP/PS study the absolute WTP for both programmes was required which is why the marginal WTP approach was not incorporated into the overall analysis. If the WTP values are to be used for the purpose of comparison with costs for the overall cost-benefit analysis of programmes then the elicitation of the relative WTP between programmes is not that informative. Analysts need to know what the overall preference is for the programme to compare against the resource costs hence the absolute value is needed, not the relative value.

10.16 Comparison with alternative outcome measures

The most common criticism of the WTP method is the attachment of a monetary value onto health care benefits. When operating with a limited health care budget though, it is impossible to move away from attaching monetary values to health care when making decisions [Phelps, 1991 #21]. A cost-utility analysis (CUA) and cost-effectiveness analysis (CEA) aim to maximise Quality Adjusted Life Years (QALYs) or life years saved for a given budget. The cost per life year/QALY ratio produced from health care programmes are then implemented from the most cost-effective to the least cost-effective until the budget runs out. Often the incremental cost-effectiveness ratio is considered and the decision-maker has to decide if the extra benefits produced from the health care programme are worth the extra cost (the south-west and north-east quadrants of figure 3.1, chapter three). Given that we operate in a world of scarce resources, the consideration of funding various health care programmes requires a cut-off point (sometimes referred to as λ). This is the agreed cut-off point beyond which the project is deemed not cost-effective. This means that decision-makers are always going to have to decide at what point the programme is no longer beneficial and in doing so, they are implicitly choosing a WTP for a life gained or a QALY. Therefore with any budget constraint the decision-makers are faced with deciding on the maximum acceptable level of cost-effectiveness. Instead of doing it implicitly with a CEA/CUA, a CBA explicitly defines the WTP for the benefits of the programme being evaluated. When costs and effects are measured in monetary terms, there are no need for ratios. If the difference between the effects and costs are positive the programme is implemented, if it is negative, then it is not.

What this thesis has shown is that the method by which you elicit WTP values impacts upon the final WTP results. From reading this outcome one might draw conclusions that the WTP method is not a valid measure of benefits within health care. It is important however to point out to the reader that the alternative measures of outcome, i.e. QALY measurement has also got its own shortcomings. QALYs are constructed by measuring health-related quality of life on a cardinal scale between 0 (death) and 1 (full health). The QALY weights of different health states are often referred to as health-state utilities.

There are three different main methods for measuring health-state utilities on a cardinal scale: the rating scale, the time trade-off and the standard gamble [Torrance, 1986 #22]. A few comparative studies have found these methods of measuring health-state utilities to produce different results [Torrance, 1976 #23][Wolfson, 1982 #24][Quinn, 1981 #25]. Read et al. compared the standard gamble, time trade-off and category scaling method for assessing preferences among hypothetical outcomes of coronary artery bypass surgery and found that each method produced different scale values. The authors suggest that different assessment methods lead people to construct different preferences and that it may not be valid to consider one method of preference assessment as a standard for another [Read, 1984 #26]. Therefore if these alternative methods of health outcome measurement contain the same amount of discrepancy between the different ways in which a question can be 'framed' then it follows that perhaps the WTP method is not that different. This thesis has shown that the different elicitation designs produce different WTP values, not unlike the empirical work that has shown the standard gamble, time-trade off and category scaling to produce different health state utilities.

10.17 Conclusions

Although the work conducted for this thesis has not answered all the questions around the methodology of the WTP method, it is hoped that it has contributed and helped to move forward the debate around the appropriate elicitation format. Inevitably future research questions have arisen from doing the work and these (hopefully) will be part of my future work in the area. It is evident however from using the WTP method that individuals place value on aspects of a health care programme that are not captured by the alternative outcome measures such as 'life years gained' and 'QALYs'. There is still a lot of work to be done in refining the methodology but the method does have a lot of advantages and it is important to remember that there is no existing alternative method of measuring the sort of benefits that are important to individuals. If we want to move forward in incorporating patient values into our decision making then methods such as the WTP method that provide a broader perspective have to be used.

References

Altman, D. G. (1991). Preparing to analyse data. Practical Statistics for Medical Research. London, Chapman & Hall.

Andreoni, J. (1989). "Giving with impure altruism: applications to charity and ricardian equivalence." Journal of Political Economy **97**: 1447-1458.

Andreoni, J. (1990). "Impure altruism and donations to public goods: A theory of warm-glow giving." Economic Journal **100**: 464-477.

Arrow, K. J. (1971). Uncertainty and the welfare economics of medical care. Essays in the theory of risk-bearing. K. J. Arrow. Chicago, Markham Publishing Company: 177.

Arrow, K., R. Solow, et al. (1993). Report of the NOAA Panel of Contingent Valuation. Federal Register. Washington DC. 58: 4601-4614.

Atkin, W. S., A. Hart, et al. (1998). "Uptake, yield of neoplasia, and adverse effects of flexible sigmoidoscopy screening." Gut **42**: 560-565.

Badia, X., M. Roset, et al. (1999). "Inconsistent responses in three preference-elicitation methods for health states." Social Science and Medicine **49**: 943-950.

Bala, M.V., Wood, L.L., Zarkin, G.A., Norton, E.C., Gafni, A., O'Brien, B. (1998) "Valuing outcomes in health care: a comparison of willingness to pay and quality-adjusted life-years". Journal of Clinical Epidemiology **51**(8): 667-76.

Barner, J.C., Mason, H.L., Murray, M.D. (1999) "Assessment of asthma patients' willingness to pay for and give time to an asthma self-management program". Clinical Therapeutics **21**(5):878-94.

Baron, J. and N. P. Maxwell (1996). "Cost of Public Goods Affects Willingness to Pay for Them." Journal of Behavioural Decision Making **9**: 173-183.

Bassett, G.W. (1997). The St. Petersburg Paradox and Bounded Utility In: Expected utility, fair gambles and rational choice. Cheltenham, Elgar Reference Collection: 181-187

Bateman, I., K. Willis, et al. ESRC Countryside Change Initiative. Consistency between CV estimates: a comparison of 2 studies of UK National Parks.: 1-44.

Bateman, I.J., Langford, I.H., Munro, A., Starmer, C., Sugden, R. (2000) " Estimating Four Hicksian Welfare Measures for a Public Good: A Contingent Valuation Investigation". Land Economics **76**(3): 355-373.

Berwick, D. and M. Weinstein (1985). "What do patients value? Willingness to pay for ultrasound in normal pregnancy." Medical Care **23**: 881-893.

Bishai, D.M., Lang, H.C. (2000) "The willingness to pay for wait reduction: the disutility of queues for cataract surgery in Canada, Denmark, and Spain". Journal of Health Economics **19**(2): 219-30.

Bishop, R., P. Champ, et al. (1995). Contingent Valuation. The Handbook of Environmental Economics. D. Bromley. Massachusets, Blackwell Publishers Ltd: 629-654.

Bishop, R., T. Heberlein, et al. (1983). "Contingent valuation of environmental assets: comparisons with a simulated market." Natural Resources Journal **23**: 619-633.

Bishop, R. C. and T. A. Heberlein (1979). "Measuring Values of Extra Market Goods: Are Indirect Measures Biased?" American Journal of Agricultural Economics **61**(5): 926-930.

Blumenschein, K., Johannesson, M. (1998) "Relationship between quality of life instruments, health state utilities, and willingness to pay in patients with asthma." Annals of Allergy, Asthma, & Immunology **80**(2): 189-94.

Blumenschein, K., Johannesson, M., Yokoyama, K.K., Freeman, P.R. (2001) "Hypothetical versus real willingness to pay in the health care sector: results from a field experiment." Journal of Health Economics **20**(3): 441-57.

Bohm, P. (1972). "Estimating Demand for Public Goods: An Experiment." European Economic Review **3**: 111-130.

Boyle, K., R. Bishop, et al. (1985). "Starting point bias in contingent valuation bidding games." Land Economics **61**: 188-194.

Brooks, A. (1998). "Pilot bowel screening scheme to start in Britain." British Medical Journal **317**: 769.

Brookshire, D. and D. Coursey (1987). "Measuring the value of a public food: an empirical comparison of elicitation procedures." American Economic Review 77(4): 554-566.

Brookshire, D., R. d'Arge, et al. (1981). Experiments in valuing public goods. Advances in Applied Microeconomics. V. Smith. Connecticut, JAI Press Inc. 1.

Brookshire, D., A. Randall, et al. (1980). "Valuing increments and decrements of natural resource service flows." American Journal of Agricultural Economics **62**: 478-488.

Brown, T. C., P. A. Champ, et al. (1996). "Which response format reveals the truth about donations to a public good?" Land Economics **71**: 152-166.

Cameron, T. A. (1988). "A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression." Journal of Environmental Economics & Management **15**: 355-379.

Cancer, I. A. f. R. o. (1996). Cancer incidence, mortality and prevalence in the European Union, EUCAN 1996 estimates. (www.dep.iarc.fr/eucan/eucan.htm). 2002.

Carson, R. T. and et al (1994). A Bibliography of Contingent Valuation Studies and Papers. La Jolla CA, Natural Resource Damage Assessment Inc.

Carson, R. T., W. M. Hanemann, et al. (1986). Determining the Demand for Public Goods by Simulating Referendums at Different Tax Prices. manuscript, University of California, San Diego.

Chestnut, L., L. Keller, et al. (1996). "Willingness to pay for changes in angina symptoms." Medical Decision Making **16**: 248-253.

Cho-Min-Naing, S. Lertmaharit, et al. (2000). "Ex post and ex ante willingness to pay (WTP) for the ICT malaria kit in Myanmar." Southeast Asian Journal of Tropical Medicine and Public Health **31**(1): 104-111.

Clarke, P. M. (2002). "Testing the convergent validity of the contingent valuation and travel cost methods in valuing the benefits of health care." Health Economics **11**: 117-127.

Clawson, M. (1959). Methods of Measuring the Demand for and the Value of Outdoor Recreation. Washington D.C., Resources for the Future.

Clawson, M. and J. Knetsch (1966). Economics of Outdoor Recreation. Baltimore, The John Hopkins University Press for Resources for the Future.

Cleary, P. D. (1999). "The increasing importance of patient surveys. Now that sound methods exit, patient surveys can facilitate improvements." British Medical Journal **319**: 720-721.

Cleary, P. D. and B. J. McNeil (1988). "Patient satisfaction as an indicator of quality care." Inquiry **25**: 25-36.

Coley, C.M., Li, Y-H., Medsger, A.R., Marrie, T.J., Fine, M.J., Kapoor, W.N., Lave, J.R., Detsky, A.S., Weinstein, M.C., Singer, D.E. (1996). "Preferences for home vs. hospital care among low-risk patients with community-acquired pneumonia." Archives of Internal Medicine **156**: 1565-1571.

Corso, P.S., Hammitt, J.K., Graham, J.D., Dicker, R.C., Goldie, S.J. (2002). "Assessing preferences for prevention versus treatment using willingness to pay." Medical Decision Making **22**(5 Suppl): S92-101.

Coursey, D., J. Hovis, et al. (1987). "The disparity between willingness to accept and willingness to pay measures of value." Quarterly Journal of Economics **102**: 679-690.

Cubitt, R., A. Munro, et al. (2000). Preference Reversal: An Experiment Investigation of Economic and Psychological Hypotheses. School of Economics, University of Nottingham.

Cummings, R. G., D. S. Brookshire, et al. (1986). Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Method. Totowa N.J., Rowman and Allanheld.

Dalmau-Matarrodona, E. (2001). "Alternative approaches to obtain optimal bid values in contingent valuation studies and to model protest zeros. Estimating the determinants of individuals' willingness to pay for home care services in day case surgery." Health Economics **10**: 101-118.

Davey, P., Grainger, D., MacMillan, J., Rajan, N., Aristides, M., Dobson, M. (1998) "Economic evaluation of insulin lispro versus neutral (regular) insulin therapy using a willingness-to-pay approach." Pharmacoeconomics **13**(3): 347-58.

Davis, R. K. (1964). 'The Value of Big Game Hunting in a Private Forest'. Transactions of the 29th North American Wildlife and Natural Resources Conference, Washington, D.C. Wildlife Management Institute.

Diamond, P. and J. Hausman (1994). "Contingent Valuation: Is some number better than no number?" Journal of Economic Perspectives **8**(4): 45-64.

Diamond, P. A. and J. A. Hausman (1992). On contingent valuation measurements of non-use values, MIT working paper. Massachusetts.

Dickie, M. and S. Gerking (1996). "Formation of risk beliefs, joint production and willingness to pay to avoid skin cancer." The Review of Economics and Statistics **78**: 451-463.

Diener, A., B. O'Brien, et al. (1998). "Health Care Contingent Valuation Studies: A Review and Classification of The Literature." Health Economics **7**: 313-326.

Dixon, S. and P. Shackley (1998). Using contingent valuation to elicit public preferences for water fluoridation. UK Health Economics Study Group, Galway.

Dolan, P. and P. Kind (1996). "Inconsistency and health state valuations." Social Science and Medicine **42**: 609-615.

Donaldson, C. (1999). "Valuing the benefits of publicly-provided health care: does 'ability to pay' preclude the use of willingness to pay?" Social Science & Medicine **49**: 551-563.

Donaldson, C. (1995). Willingness To Pay for Publicly-Provided Health Care. Department of Economics. Aberdeen, University of Aberdeen: 9.

Donaldson, C., S. Birch, et al. (2002). "The distribution problem in economic evaluation: income and the valuation of costs and consequences of health care programmes." Health Economics **11**: 55-70.

Donaldson, C., S. Farrar, et al. (1997). "Assessing community values in health care: is the willingness to pay method feasible?" Health Care Analysis **5**: 7-29.

Donaldson, C. and Gerard, K. (1993). Economics of health care financing: the visible hand. London, MacMillan.

Donaldson, C., V. Hundley, et al. (1998). "Willingness to pay: a method for measuring preferences for maternity care?" BIRTH **25**(1): 32-39.

Donaldson, C., A. M. Jones, et al. (1998). "Limited dependent variables in willingness to pay studies: applications in health care." Applied Economics **30**: 667-677.

Donaldson, C. and P. Shackley (1997). "Does 'process utility' exist? A case study of willingness to pay for laparoscopic cholecystectomy." Social Science & Medicine **44**(5): 699-707.

Donaldson, C., P. Shackley, et al. (1997). "Using willingness to pay to value close substitutes: carrier screening for cystic fibrosis revisited." Health Economics **6**: 145-159.

Donaldson, C., R. Thomas, et al. (1995). Open-ended versus payment scale approaches to eliciting willingness to pay. Health Economics Research Unit Discussion, Aberdeen.

Dranitsaris,G. (1999) "Pamidronate for the prevention of skeletal-related events in multiple myeloma. What does the public think it is worth?" International Journal of Technology Assessment in Health Care **15**(1): 108-22.

Dranitsaris, G., Longo, C.J., Grossman, L.D. (2000) "The economic value of a new insulin preparation, Humalog Mix 25. Measured by a willingness-to-pay approach". Pharmacoeconomics **18**(3): 275-87.

Drummond, M. F., B. O'Brien, et al. (1997). Methods for the Economic Evaluation of Health Care Programmes. New York, Oxford University Press.

Dubourg, W., M. Jones-Lee, et al. (1994). "Imprecise preferences and the WTP-WTA disparity." Journal of Risk and Uncertainty **9**: 115-133.

Eastaugh, S. (1991). "Valuation of the Benefits of Risk-Free Blood. Willingness to pay for hemoglobin solutions." International Journal of Technology Assessment in Health Care **7**(1): 51-57.

Eastaugh, S. R. (2000). "Willingness to pay in treatment of bleeding disorders." International Journal of Technology Assessment in Health Care **16**(2): 706-710.

Evans, R., M. Barer, et al. (1994). "Why are Some People Healthy and Others Not?" The Determinants of Health of Populations. New York, Aldine de Gruyter.

Fischhoff, B., Bostrom, A., Quadrel, M.J. (1993). "Risk perception and communication." Annual Review of Public Health **14**: 183-203.

Forsythe, S., Arthur, G., Ngatia, G., Mutemi, R., Okhaimbo, J., Gilks, C. (2002). "Assessing the cost and willingness to pay for voluntary HIV counselling and testing in Kenya." Health Policy & Planning **12**(2): 187-95.

Frew, E. J., V. Hammersley, et al. (2001). "Collaborating with a primary care-based research network." Journal of Evaluation in Clinical Practice **7**(3): 339-342.

Frew, E., J. L. Wolstenholme, et al. (1999). "Estimating time and travel costs incurred in clinic-based screening: flexible sigmoidoscopy screening for colorectal cancer." Journal of Medical Screening **6**: 119-123.

Frew, E., J. L. Wolstenholme, et al (2001). "Willingness to pay for colorectal cancer screening." European Journal of Cancer **37**: 1746-1751.

Garbacz, C. and M. Thayer (1983). "An experiment in valuing senior companion program services." Journal of Human Resources **18**(1): 147-153.

Gerard, A., M. Buyse, et al. (1988). "Preoperative radiotherapy as adjuvant treatment in rectal cancer. Final results of a randomised study (EORTC)." Annals of Surgery **208**: 606.

Gibb, S., C. Donaldson, et al. (1998). "Assessing strength of preference for abortion method using 'willingness to pay': a useful research technique for measuring values." Journal of Advanced Nursing **27**(1): 30-36.

Griggs, J. L. and A. I. Mushlin (1998). "Economic analysis of expensive technologies. The case of erythropoietin." CANCER **83**(12): 2427-2429.

Grimes, D. S. (1988). "Value of a negative cervical smear." British Medical Journal **296**: 1363.

Gastrointestinal Tumour Study Group. (1984). "Adjuvant therapy of colon cancer: results of a prospectively randomised trial." New England Journal of Medicine **311**: 310-737.

Gyrd-Hansen, D., J. Sogaard, et al. (1998). "Colorectal cancer screening: efficiency and effectiveness." Health Economics **7**: 9-20.

Gold, M., J. Siegel, et al. (1996). Cost-effectiveness in Health and Medicine. New York, Oxford University Press.

Greene, W. (1997). Econometric Analysis, Upper Saddle River: Prentice Hall International.

Hanemann, W. M. (1993). Contingent valuation and economics, Working Paper No. 697. California, Division of Agriculture and Natural Resources, University of California.

Hall, J. and M. Dornan (1988). "What patients like about their medical care and how often they are asked. A meta-analysis of the satisfaction literature." Social Science and Medicine **22**: 935-939.

Hanemann, W. M. (1994). "Valuing the Environment through Contingent Valuation." Journal of Economic Perspectives **8**(4): 19-43.

Hanemann, W. M. (1992). Preface. Pricing the European Environment. S. Navrud. Oslo, Scandinavian University Press.

Hannemann, W. (1991). "Willingness to pay and willingness to accept: how much can they differ?" American Economic Review **81**: 635-647.

Hanemann, W. and B. Kanninen (1996). The statistical analysis of discrete-response contingent valuation data, University of California at Berkeley: Department of Agriculture and Resource Economics Working Paper: 798.

Hardcastle, J. D. and T. W. Balfour (1980). "Screening for symptomless colorectal cancer by testing for occult blood in general practice." The Lancet April **12**: 791-793.

Hardcastle, J. D., J. O. Chamberlain, et al. (1996). "Randomised controlled trial of faecal-occult-blood screening for colorectal cancer." The Lancet **348**: 1472-1477.

Harless, D.W. and Camerer, C.F. (1994). "The predictive utility of generalized expected utility theories" **62**(6): 1251-1289.

Harris, B. (1984). "Contingent valuation of water pollution control." Journal of Environmental Management **19**: 199-208.

Hassan, E.O., El-Nahal, N., El-Hussein. (1994). "Acceptability of the once-a-month injectable contraceptives Cyclofem and Mesigyna in Egypt." Contraception **49**: 469-488.

Heap, S. H., M. Hollis, et al. (1992). The Theory of Choice; A Critical Guide. Oxford UK, Blackwell Publishers.

Hennekens, C. H. and J. E. Buring (1987). Epidemiology in Medicine. Boston/Toronto, Little, Brown and Company.

Herriges, J. A. and J. F. Shogren (1996). "Starting point bias in dichotomous choice." Journal of Environmental Economics and Management **30**: 112-131.

Hirth, R.A., Bloom, B.S., Chernew, M.E., Fendrick, A.M. (2000) "Patient, physician, and payer perceptions and misperceptions of willingness to pay for diagnostic certainty." International Journal of Technology Assessment in Health Care **16**(1): 35-49.

Hoehn, J. P. and A. Randall (1983). Incentives and Performance in Contingent Policy Valuation. American Agricultural Economics Association, Purdue University.

Hoehn, J. P. and A. Randall (1987). "A satisfactory benefit cost indicator from contingent valuation." Journal of Environmental Economics & Management **14**: 226-247.

Hoosein, N. M., P. A. Kiener, et al. (1988). "Antiproliferative effects of gastrin receptor antagonists and antibodies to gastrin on human colon carcinoma cell lines." Cancer Research **48**: 7179-7183.

Johannesson, M. (1992). "Economic evaluation of lipid lowering - a feasibility test of the contingent valuation approach." Health Policy **20**: 309-320.

Johannesson, M. (1996). "A note on the relationship between ex ante and expected willingness to pay for health care." Social Science & Medicine **42**(3): 305-311.

Johannesson, M. and B. Jonsson (1991). "Economic evaluation in health care: Is there a role for cost-benefit analysis?" Health Policy **17**: 1-23.

Johannesson, M., B. Jonsson, et al. (1991). "Willingness to pay for antihypertensive therapy: Results of a Swedish pilot study." Journal of Health Economics **10**: 461-474.

Johannesson, M. and D. Meltzer (1998). "Some reflections on cost-effectiveness analysis." Health Economics **7**: 1-7.

Johannesson, M., R. M. O'Conor, et al. (1997). "Willingness to pay for reduced incontinence symptoms." British Journal of Urology **80**: 557-562.

Johnson, R., N. Bergenzer, et al. (1990). Contingent valuation question formats: dichotomous choice versus open-ended responses. Economic Valuation of Natural Resources: Issues, Theory and Applications. R. Johnson and G. Johnson. Boulder, Westview Press: 193-204.

Kahneman, D. and J. L. Knetsch (1992). "Valuing Public Goods: The Purchase of Moral Satisfaction." Journal of Environmental Economics and Management **22**: 57-70.

Kartman, B., F. Andersson, et al. (1996). "Willingness to pay for reductions in angina pectoris attacks." Medical Decision Making **16**: 248-253.

Kartman, B., N. Stalhammar, et al. (1996). "Valuation of health changes with the contingent valuation method: A test of scope and question order effects." Health Economics **5**: 531-541.

Kartman, B., N. Stalhammar, et al. (1997). "Contingent valuation with an open-ended follow-up question: a test of scope effects." Health Economics **6**: 637-639.

Kealy, M., M. Montgomery, et al. (1990). "Reliability and predictive validity of contingent values: does the nature of the good matter?" Journal of Environmental Economics and Management **19**: 244-263.

Kealy, M. J. and R. W. Turner (1993). "A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuations." American Journal of Agricultural Economics **75**: 321-331.

Keith, P., J. Haddon, et al. (2000). "A cost-benefit analysis using a willingness to pay questionnaire of intranasal budesonide for seasonal allergic rhinitis." Annals of allergy, asthma & immunology **84**: 55-62.

Kemp, M. A. and Maxwell (1993). Exploring a budget context for contingent valuation estimates. Contingent valuation: a critical assessment. J. Hausman. Amsterdam, North Holland Press: 217-270.

Kenkel, D. (1997). "On valuing morbidity, cost-effectiveness analysis and being rude." Journal of Health Economics **16**: 749-758.

Kidholm, K. and L. K. Langkilde (1994). "Interpretation of Zero Bids in Contingent Valuation Surveys." Paper presented at the 15th Nordic Health Economist's Study Group Meeting in Reykjavik.: 1-25.

Klose, T. (1999). "The contingent valuation method in health care." Health Policy **47**: 97-123.

Kneese, A. V. (1966). Research Goals and Progress toward Them. Environmental Quality in a Growing Economy. H. Jarrett. Washington, Johns Hopkins Press.

Kneese, A. (1984). Measuring the Benefits of Clean Air and Water. Washington, D.C., Resources for the Future.

Knetsch, J. and J. Sinden (1984). "Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value." Quarterly Journal of Economics **99**(3): 507-521.

Krabbe, P. F. M., M. L. Essink-Bot, et al. (1997). "The comparability and reliability of five health-state valuation methods." Social Science and Medicine **45**: 1641-1652.

Kronborg, O., C. Fenger, et al. (1996). "Randomised study of screening for colorectal cancer with faecal-occult-blood test." The Lancet **348**: 1467-1471.

Kruitilla, J. V. and A. C. Fisher (1975). The Economics of Natural Environments: Studies in the Valuation of Commodity and Amenity Resources. Baltimore, The John Hopkins University Press for Resources for the Future.

Kuppermann, M., Nease, R.F. Jr., Ackerson, L.M., Black, S.B., Shinefield, H.R., Lieu, T.A. (2000) "Parents' preferences for outcomes associated with childhood vaccinations." Paediatric Infectious Disease Journal **19**(2):129-33.

Langford, I. H., I. J. Bateman, et al. (1996). "A multilevel modelling approach to triple bounded dichotomous choice contingent valuation." Environmental and Resource Economics **7**: 197-211.

Larson, R.A. (2000) "Patients' willingness to pay for pharmaceutical care." Journal of the American Pharmaceutical Association **40**(5): 618-24.

Lee, S. J., B. Liljas, et al. (1998). "The impact of risk information on patients' willingness to pay for autologous blood donation." Medical Care **36**(8): 1162-1173.

Lewis, J. D. and D. A. Asch (1999). "Barriers to office-based screening sigmoidoscopy: does reimbursement cover costs?" Annals of Internal Medicine **130**: 525-530.

Lichtenstein, S. and P. Slovic (1971). "Reversals of preference between bids and choices in gambling decisions." Journal of Experimental Psychology **89**: 49-55.

Lindholm, L., M. Rosen, et al. (1994). "Are people willing to pay for a community-based preventative program." International Journal of Technology Assessment in Health Care **10**: 317-324.

Lindholm, L.A., Rosen, M.E., Stenbeck, M.E. (1997) " Determinants of willingness to pay taxes for a community prevention programme." Scandinavian Journal of Social Medicine **25** (2): 126-135.

Lindman, H. (1971). "Inconsistent preferences among gambles." Journal of Experimental Psychology **80**: 590-597.

Liu, J.T., Hammitt, J.K., Wang, J.D., Liu, J.L. (2000) "Mother's willingness to pay for her own and her child's health: a contingent valuation study in Taiwan." Health Economics. **9**(4): 319-26.

Loomes, G., Starmer, C., Sugden, R. (1989). "Preference reversal: Information-processing effect or rational non-transitive choice?" The Economic Journal **99** (395):140-151.

Loomis, J., T. Brown, et al. (1997). "Evaluating the validity of the dichotomous choice question format in contingent valuation." Environmental and Resource Economics **10**: 109-123.

Luchini, S., Protiere, C., Moatti, J-P. (2002). "Eliciting several willingness to pay in a single contingent valuation study: application to health care." Health Economics **12**: 51-64.

Lundberg, L., Johannesson, M., Silverdahl, M., Hermansson, C., Lindberg, M. (1999) "Quality of life, health-state utilities and willingness to pay in patients with psoriasis and atopic eczema." British Journal of Dermatology **141**(6): 1067-1075.

Marteau, T. M. (1989). "Psychological costs of screening. May sometimes be bad enough to undermine the benefits of screening." British Medical Journal **299**: 527.

Matthews, D.,C. Birch, S., Gafni, A., DiCenso, A. "Willingness to pay for periodontal therapy: development and testing of an instrument." Journal of Public Health Dentistry **59**(1): 44-51

Mayhew, G. and R. Winer (1992). "An empirical analysis of internal and external reference prices using scanner data." Journal of Consumer Research **19**: 62-70.

Miedzybrodzka, Z., Semper, J., Shackley, P., Abdalla, M., Donaldson, C. (1995). "Stepwise or couple antenatal carrier screening for cystic fibrosis?: women's preferences and willingness to pay." Journal of Medical Genetics **32**: 282-283.

Milon, J. (1989). "Contingent Valuation Experiments for Strategic Behaviour." Journal of Environmental Economics and Management **17**: 293-308.

Mishan, E. (1971). "Evaluation of life and limb: a theoretical approach." Journal of Political Economy **79**: 687-705.

Mitchell, R. and R. Carson (1989). Using surveys to value public goods: the contingent valuation method. Washington, DC, Resources for the future.

Mitchell, R. C. and R. T. Carson (1981). An Experiment in Determining Willingness to Pay for National Water Quality Improvements. Washington D.C., U.S. Environmental Protection Agency.

Mitchell, R. C. and R. T. Carson (1984). A Contingent Valuation Estimate of National Freshwater Benefits: Technical Report to the U.S. Environmental Protection Agency. Washington D.C., Resources for the Future.

Mooney, G. (1994). Key issues in health economics. London, Harvester Wheatsheaf.

Morrison, G. C. (1997) "Willingness to pay and willingness to accept: some evidence of an endowment effect." Applied Economics **29**: 411-417.

Muller, A. and T. J. Reutzel (1984). "Willingness to pay for reduction in fatality risk: an exploratory study." American Journal of Public Health **74**(8): 808-812.

Muto, T., H. J. R. Bussey, et al. (1975). "The evolution of cancer of the colon and rectum." Cancer **36**: 2251-2270.

Narbro, K. and L. Sjostrom (2000). "Willingness to pay for obesity treatment." International Journal of Health Technology Assessment **16**(1): 50-59.

Neilson, A. R. and D. K. Whynes (1995). "Determinants of persistent compliance with screening for colorectal cancer." Social Science & Medicine **41**: 365-374.

Neumann, P. and M. Johannesson (1994). "The willingness to pay for in vitro fertilisation: a pilot study using contingent valuation." Medical Care **32**: 686.

Neumann, J.V. and Morgenstern, O. (1944) Theory of Games and Economic Behavior. Princeton University Press.

Newcomb, P. A., R. G. Norfleet, et al. (1992). "Screening sigmoidoscopy and colorectal cancer mortality." Journal of the National Cancer Institute **84**: 1572-1575.

Nivatvongs, S. and D. S. Fryd (1980). "How far does the proctosigmoidoscope reach? A prospective study of 1000 patients." New England Journal of Medicine **303**(7): 380-382.

Nord, E. (1992). "Methods for Quality Adjustment of Life Years." Social Science and Medicine **34**(5): 559-569.

O'Brien, B. and A. Gafni (1996). "When do the 'dollars' make sense? Toward a conceptual framework for contingent valuation studies in health care." Medical Decision Making **16**: 288.

O'Brien, B., R. Goeree, et al. (1998). "Assessing the Value of a New Pharmaceutical. A feasibility study of contingent valuation in managed care." Medical Care **36**(3): 370-384.

O'Brien, B. and J. Viramontes (1994). "Willingness to pay: a valid and reliable measure of health state preference?" Medical Decision Making **14**: 289.

Office of National Statistics (1995-1998). Distribution of Household Income : Regional Trends Dataset RT34802, www.statistics.gov.uk/statbase/xsdataset.asp. 2000.

Oliver, A. (2002). "Complete preferences over health states": reply to the paper by Shiell et al. Health Economics **9**(8): 727-728.

Olsen, J. (2000). Willingness to pay: Experience with their use in economic evaluation. European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium.

Olsen, J. A. and C. Donaldson (1998). "Helicopters, hearts and hips: using willingness to pay to set priorities for public sector health care programmes." Social Science and Medicine **46**(1): 1-12.

Office of National Statistics (1995-1998). Family Expenditure Survey.

Onwujekwe, O. (2001) "Searching for a better willingness to pay elicitation method in rural Nigeria: the binary question with follow-up method versus the bidding game technique." Health Economics **10**(2):147-58.

Office for Population Census' and Surveys (1991-1992). "Morbidity Statistics from General Practice, 4th National Study."

Office for Population Census' and Surveys (1994). Cancer Statistics Registrations 1989. London, HMSO.

World Health Organisation (1997). Cancer statistics (www.who.int).

Ortega, A., Dranitsaris, G., Puodziunas, A. L. (1998) "What are cancer patients willing to pay for prophylactic epoetin alfa? A cost-benefit analysis." Cancer **83**(12): 2588-2596

Osmond, M.H., Klassen, T.P., Quinn, J.V. (1995). "Economic comparison of a tissue adhesive and suturing in the repair of paediatric facial lacerations." Journal of Paediatrics **126**: 892-895.

Papatheofanis, F.J. (2000) "The willingness to pay for positron emission tomography (PET): evaluation of suspected lung cancer using contingent valuation." Quarterly Journal of Nuclear Medicine **44**(2):191-6.

Pauly, M. V. (1996). Valuing health care benefits in money terms. Valuing Health Care. Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. F. A. Sloan. Cambridge, Cambridge University Press.

Pearce, D. W. (1983). Cost-Benefit Analysis. London, Macmillan Education Ltd.

Pearce, D. W. and C. A. Nash (1981). The Social Appraisal of Projects A Test in Cost-Benefit Analysis. London, MacMillan Press Ltd.

Phelps, C. and A. I. Mushlin (1991). "On the near equivalence of cost-effectiveness analysis." International Journal of Technology Assessment in Health Care **17**(1): 12-21.

Poyner, T.F., Menday, A.P., Williams, Z.V. (2000) "Patient attitudes to topical antipsoriatic treatment with calcipotriol and dithranol." Journal of the European Academy of Dermatology & Venereology. **14**(3): 153-158.

Quinn, R. J. (1981). "The effect of measurement on preference scales in a medical decision making context." Medical Decision Making **1**: 431.

Rajendran, K. N. and G. J. Tellis (1994). "Contextual and temporal components of reference price." Journal of Marketing **58**: 22-34.

Ramsey, S. D., S. D. Sullivan, et al. (1997). "Willingness To Pay for Antihypertensive Care: Evidence from a Staff-Model HMO." Social Science & Medicine **44**(12): 1911-1917.

Randall, A., B. Ives, et al. (1974). "Bidding games for valuation of aesthetic environmental improvements." Journal of Environmental Economics and Management **1**: 132-149.

Read, J. L., Quinn, R. J., Berwick, D. M., Fineberg, H.V., Weinstein, M.C. (1984). "Preferences for Health Outcomes. Comparison of Assessment Methods." Medical Decision Making **4**(3): 315-329.

Reardon, G. and D. S. Pathak (1989). "Contingent valuation of pharmaceuticals and pharmacy services - methodological considerations." Journal of Social and Administrative Pharmacy **6**(2): 83-91.

Reis, J., Sherman, S., Macon, J., Friedman, B. (1990). " Care for the Underinsured: Who Should Pay?" Journal of Nursing Administration **20**(3): 16-20.

Ridker, R. G. (1967). The Economic Costs of Air Pollution. New York, Praeger.

Robine, J. and K. Ritchie (1991). "Healthy life expectancy: Evaluation of global indicator of change in population health." British Medical Journal **302**: 457-460.

Rosen, S. (1974). "Hedonic prices and implicit markets: product differentiation in pure competition." Journal of Political Economy **82**: 34-55.

Rowe, R., R. d'Arge, et al. (1980). "An experiment on the economic value of visibility." Journal of Environmental Economics and Management **7**: 1-19.

Roth, A. E. (1995). Introduction to experimental economics. The handbook of experimental economics. J. H. Kagel and A. E. Roth. Princeton NJ, Princeton University Press: 3-109.

Ryan, M. (1996). "Using willingness to pay to assess the benefits of assisted reproductive techniques." Health Economics **5**: 543-558.

Ryan, M. and F. S. Miguel (2000). "Testing for consistency in willingness to pay experiments." Journal of Economic Psychology **21**: 305-317.

Ryan, M., J. Ratcliffe, et al. (1997). "Using willingness to pay to value alternative modes of antenatal care." Social Science and Medicine **44**: 371-380.

Schkade, D. and J. Payne (1993). Where do the numbers come from? How people respond to contingent valuation questions. Contingent Valuation: A Critical Assessment. J. Hausman. Amsterdam, North Holland Press: 271-304.

Schkade, D. A. and J. W. Payne (1994). "How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation." Journal of Environmental Economics and Management **26**: 88-109.

Selby, J. V., G. D. Friedman, et al. (1992). "A case-control study of screening sigmoidoscopy and mortality from colorectal cancer." New England Journal of Medicine **326**: 653-657.

Seller, C., J. R. Stoll, et al. (1985). "Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques." Land Economics **61**: 156-175.

Sen, A. (1993). "Internal consistency of choice." Econometrica **61**: 495-521.

Shackley, P., Donaldson, C. (2002) "Should we use willingness to pay to elicit community preferences for health care? New evidence from using a 'marginal' approach." Journal of Health Economics **21**(6): 971-91.

Shiell, A., Seymour, J., Hawe, P., Cameron, S. (2000) "Are preferences over health states complete?" Health Economics **9**(1): 47-55.

Shiell, A., Hawe, P., Seymour, J. (2000) "Are preferences over health states complete?" [comment] Health Economics **9**(8): 729-731.

Shiell, A., Gold, L. (2003) "If the price is right: vagueness and values clarification in contingent valuation." Health Economics: in press.

Slothuus, U. and R. Brooks (2000). "Willingness to pay in arthritis: a Danish contribution." Rheumatology **39**: 791-799.

Slothuus, U., Larsen, M., Peter, J. (2000). "Willingness to pay for arthritis symptom alleviation. Comparison of closed-ended questions with and without follow-up." International Journal of Technology Assessment in Health Care **16**(1): 60-72.

Smith, D. (1983). "APhA national survey: willingness of consumers to pay for pharmacists' clinical services." Annals of Pharmacy **NS23**(6): 58-64.

Smith, R. D. (2000). "The discrete-choice willingness to pay question format in health economics: should we adopt environmental guidelines?" Medical Decision Making **20**: 194-206.

Smith, R.D. (2001) "The relative sensitivity of willingness-to-pay and time-trade-off to changes in health status: an empirical investigation." Health Economics **10**(6): 487-97.

Smith, V. L. (1979). "An Experimental Comparison of Three Public Good Decision Mechanisms." Scandinavian Journal of Economics **81**: 198-215.

Sorum, P.C. (1999) "Measuring patient preferences by willingness to pay to avoid: the case of acute otitis media." Medical Decision Making **19**(1): 27-37.

Stalhammer, N. (1996). "An empirical note on willingness to pay and starting point bias." Medical Decision Making **16**: 242-247.

Starmer, C. (2000). "Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk." Journal of Economic Literature **38**: 332-382.

Stephens, M.B., Montefalcon, R., Lane, D.A. (2000) "The maternal perspective on prenatal ultrasound." Journal of Family Practice **49**(7): 601-604.

Stewart, J. M., O'Shea, E., Donaldson, C., Shackley, P. (2002). "Do ordering effects matter in willingness to pay studies of health care?" Journal of Health Economics **21**: 585-599.

Stryker, S. J., B. G. Wolff, et al. (1987). "Natural history of untreated colonic polyps." Gastroenterology **93**: 1009-1013.

Sutton, S., J. Wardle, et al. (2000). "Predictors of attendance in the United Kingdom flexible sigmoidoscopy screening trial." Journal of Medical Screening **7**: 99-104.

Swan, J., D. Fryback, et al. (1997). "MR and conventional angiography: work in progress towards assess utility in radiology." Academic Radiology **4**: 475-482.

Taylor, S.J., Armour, C.L. (2000) "Measurement of consumer preference for treatments used to induce labour: a willingness-to-pay approach." Health Expectations **3**: 203-216.

Thayer, M. (1981). "Contingent valuation techniques for assessing environmental impacts: further evidence." Journal of Environmental Economics and Management **8**: 27-44.

Thompson, M., L. Read, et al. (1984). "Feasibility of willingness-to-pay measurement in chronic arthritis." Medical Decision Making **4**: 195-215.

Tolley, G. S. and e. al. (1983). Establishing and Valuing the Effects of Improved Visibility in the Eastern United States. Report to the U.S. Environmental Protection Agency, Washington D.C.

Torrance, G.W. (1976). "Social preferences for health states. An empirical evaluation of three measurement techniques." Socio-economics Planned Science **10**:129-136.

Torrance, G.W. (1986). "Measurements of health status utilities for economic appraisal: a review." Journal of Health Economics **5**: 1-30.

Towler, B., L. Irwig, et al. (1998). "A systematic review of the effects of screening for colorectal cancer using the faecal occult blood test, Hemoccult." British Medical Journal **317**: 559-565.

Vernon, S. W. (1997). "Participation in colorectal cancer screening: a review." Journal of the National Cancer Institute **89**: 1406-1422.

Walker, A. R., D. K. Whynes, et al. (1991). "The hospital costs of diagnostic procedures for colorectal cancer." Journal of Clinical Epidemiology **44**: 907-914.

Walker, A. R., D. K. Whynes, et al. (1991). "The cost of screening of colorectal cancer." Journal of Epidemiology in Community Health **45**: 220-224.

Weaver, M., Ndamobissi, R., Kornfield, R., Blewane, C., Sathe, A., Chapko, M., Bendje, N., Nguembi, E., Senwara-Defiobona, J. (1996) "Willingness to pay for child survival: results of a national survey in central African republic." Social Science & Medicine **43**(6): 985-998.

Whynes, D. K., E. J. Frew, et al. (2002). "The costs of flexible sigmoidoscopy screening for colorectal cancer in the United Kingdom." International Journal of Health Technology Assessment: in press.

Whynes, D. K., A. R. Neilson, et al. (1998). "Faecal occult blood screening for colorectal cancer: it is cost effective?" Health Economics **7**: 21-29.

Wilking, N., N. J. Petrelli, et al. (1986). "A comparison of the 25cm rigid proctosigmoidoscope with the 65cm flexible endoscope in the screening of patients for colorectal carcinoma." Cancer **57**(3): 669-671.

Williams, A. (1985). "Economics of coronary artery bypass grafting." British Medical Journal **291**: 326-329.

Williams, C. (1986). "Who's for CO2?" Gastroenterol Endoscopy **32**(5): 365-367.

Willig, R. D. (1976). "Consumer's Surplus Without Apology." American Economic Review **66**(4): 587-597.

Wilkinson, R. (1996). Unhealthy Societies: The Afflictions of Inequality. London, Routledge.

Winawer, S. T. (1980). "Screening for colorectal cancer." Cancer **45**: 1093-1098.

Winawer, S. J., A. G. Zauber, et al. (1987). "The national polyp study: temperol sequence of evolving colorectal cancer from the normal colon (abstract)." Gastrointestinal Endoscopy **33**: A167.

Winer, R. (1986). "A reference price model of brand choice for frequently purchased products." Journal of Consumer Research **13**: 250-256.

Wolfson, A., Sinclair, A., Bombardier, C., McGeer, A., (1982). Preference measures for functional status in strobe patients. Interrater and intertechnique comparisons. In, Kane RL, Kane RA eds: Values and Long-Term Care. Lexington MA: Lexington Books.

Young, G. P., F. A. Macrae, et al. (1996). Clinical methods for early detection: basis, use and evaluation. Prevention and early detection of colorectal cancer. G. P. Young and B. Levin. London, WB Saunders Company Ltd: 242-270.

Zarkin, G.A., Cates, S.C., Bala, M.V. (2000) "Estimating the willingness to pay for drug abuse treatment: a pilot study." Journal of Substance Abuse Treatment **18**(2): 149-59.

Zethraeus, N. (1998) "Willingness to pay for hormone replacement therapy." Health Economics 7(1):31-8.

Appendix One

Table A1.1 Summary of literature search.

Author	Year	WTP/ WTA	Study Method		Sample Size	Response Rate	Brief Description
			Question Format	Collection			
Muller A and Reutzel TJ.	1984	WTP	Open ended	questionnaires	77 students	89%	The results of a survey of 77 senior year undergraduate students show that reasonable appearing and consistent responses to WTP questions on car crash protection can be obtained. The study varies the level of risk and the WTP is found to increase with the size of the risk reduction. However, the implied value of life was over 100 times greater for an unidentified life than for the respondent's own life.
Grimes D	1988	WTP	Payment scale	questionnaire	100 patients	84%	Used WTP to establish the value given by 100 women of a negative smear test result. Women gave a negative smear a mean value of £8.25. The service can therefore be considered to be cost effective as its value exceeds its cost.
Reis J et al.	1990	WTP	Open-ended	interview	150 patients	100%	Survey designed to measure inner-city medically insured adults preferences for health care services. 35% of the sample said that they would not pay a yearly fee to the Department of Health for health care services, 17% had a tentative interest and 46% had a definite interest in the system. The information obtained from this study provide information for nurse managers to help to balance patient needs against admin pressures to generate revenue for patient fees.
Johannesson M	1992	WTP	Closed-ended	Mail questionnaire	94 patients	66%	The paper reports on a pilot study of three benefit measures designed to measure patient preferences for lowering their cholesterol levels. The respondents were on average WTP about SEK 450 per month. The study set out to test the feasibility of using the WTP method in this setting and it is concluded that given its inherent advantages WTP deserves further attention.
O'Brien and	1994	WTP	Bidding game	interview	102 patients	76.6%	Comparison of the WTP method with other instruments that

Viramontes							measure health state preference, e.g. standard gamble, rating scale, SF-36. Tested for starting pt bias, convergent validity and test-retest reliability. Found no evidence of starting point bias, some evidence of convergent validity between WTP and standard gamble method and evidence of test-retest reliability with the WTP method.
Hassan et al	1994	WTP	Not specified	Interviews	1091 women		A study measuring the acceptability of two once-a-month injectable contraceptives in Egypt. Acceptibility of the two contraceptives proved to be high judged by the WTP for them when the injectables are marketed. The study provides in-depth information about the social aspects affecting the acceptability of monthly contraceptives in the different Egyptian localities.
Osmond et al.	1995	WTP	Not specified	interview	30 parents	100%	Investigation of parents' WTP for treatment of pediatric facial lacerations. A convenience sample of 30 patients within a health centre were interviewed. Parents were WTP a median of $40 for tissue adhesive and $25 for dissolving sutures if only non-dissolving sutures were provided by the health care system. Therefore it was concluded that tissue adhesive is the preferred method of closure of pediatric facial lacerations.
Donaldson et al.	1995	WTP	Open ended and Payment Scale	Mail Questionnaire	380 screening patients	61%	The open-ended and payment scale approaches to eliciting WTP are compared. The payment scale led to a higher response rate, higher rate of completion, higher mean and median WTP and a stronger association between WTP and ATP. The authors infer that the payment scale technique leads to more valid WTP values than use of the open-ended approach.
Miedzybrodzka Z et al	1995	WTP	Open-ended	questionnaire	450 women	Not reported	Study set up to measure women's preferences for antenatal screening: stepwise versus couple antenatal screening for cystic fibrosis. The mean WTP for stepwise screening was £19 and for couple screening it was £18. The WTP for stepwise was 6% greater than for couple screening – the difference is not significant.
Baron J and	1996	WTP	Open-ended	Questionnaire	3 sub-studies	Not	Authors found the judgement of WTP for the goods was

Maxwell NP					53,41,19 students.	reported	affected by cost as well as benefit, even when subjects judged the benefit to be unaffected by cost. The results can be understood as over extension of a somewhat useful heuristic: things that cost more often yield more benefit.
Kartman B et al.	1996	WTP	Closed ended/bidding game	Telephone interviews	402 patients	92%	This study measures the WTP for reductions in angina pectoris attacks. Percentage reductions in angina attacks were varied randomly in different subsamples. The WTP for a 50% reduction in attacks was lower than the 25% reduction. WTP also varied with angina pectoris status, and income in the way that the authors predicted.
Weaver M et al.	1996	WTP	Closed ended	interview	1263 households	Not reported	WTP for quality improvements in Sub-Saharan Africa was measured by CV in which subjects were asked about expenditures for care at government facilities under a hypothetical user fee/quality program. The majority of the population was WTP the cost of the quality improvements, which ranged from U.S. $0.40 to U.S. $4.00.
Ryan M	1996	WTP	Payment scale	Questionnaire	353	42%	Aim was to use the WTP technique to measure the benefit of assisted reproductive techniques other than the 'success' of leaving the service with a child. The WTP goes beyond the narrow medical definition of success and measures aspects such as the process of treatment. The results suggest that there is some value in going through the service, even if the couple leave it childless.
Coley M et al.	1996	WTP	Bidding game	interview	159 patients	50%	Measurement of patient preferences for initial outpatient versus hospital care among low risk patients being treated for pneumonia. Three utility assessment techniques were used to measure the strength of preference (category scaling, standard gamble, WTP). Of the 74% that preferred home care, there was a WTP of a mean of 24% of 1 month's household income. Most patients prefer outpatient treatment and most physicians do not include patients in decision making.
Donaldson C et al.	1997	WTP	Open ended	questionnaire	450 women attending antenatal	51%	This study involved the use of WTP to value two methods of screening for cystic fibrosis carrier status. Preferences for one of the two methods were sought prior to the WTP

					clinics.		question. Data was then analysed to reveal the overall intensity of preference for one method over the other.
M Johannesson et.al.	1997	WTP	Closed ended	questionnaire	461 patients	85%	Measured the WTP for a reduction in the number of micturitions and urinary leakages for patients with urge incontinence. The median WTP was greater for the 50% reduction than the 25% reduction. WTP was significantly related to income. Patients with incontinence problems are WTP a substantial amount for a reduction in the number of micturitions and leakages.
Ramsey et al	1997	WTP	Bidding game	questionnaires	320 patients	80%	Investigate the WT P for antihypertensive care. Participants returned a postal survey containing a WTP question with 10 'bids' ranging from \$25 to \$250. The estimated parameters were significant and indicated an adjusted mean WTP of \$93 per month for antihypertensive therapy.
Lindholm et al	1997	WTP	Bidding Technique	Interview	104	83%	Used WTP to assess the multi-dimensional consequences of prevention of disease. Found that expectations of reduced mortality and future savings in public health care spending increase the perceived value of a preventative programme. Decreased disease risk was not positively associated with WTP, while a low level of anxiety was.
Ryan M, Ratcliffe J and Tucker J.	1997	WTP	Closed-ended	Questionnaire	704	74%	Economic evaluation of alternative models of antenatal care: general practitioner/midwife routine led care versus obstetrician led care. The results suggest a WTP of £2500 for antenatal care with no significant difference between the different types of care provided.
Kartman B et al	1997	WTP	CE with follow up	Telephone interviews	402 patients	92%	Study examines the properties of open-ended questions following binary questions. Using a split sample approach, the study tests whether the open-ended follow-up is sensitive to the scope of the commodity being valued. No significant scope effects were detected. The authors conclude that the results do not support the use of a open-ended follow up question.
Morrison GC	1997	WTP/ WTA	Payment scale	Experiment	20 students	100%	An experiment was designed using students doing an experimental economics course to assess whether the endowment effect might be a significant source of the

							WTP/WTA divergence. The experiment adopts both WTP and WTA questions The study concludes that the endowment effect cannot be dismissed as a possible driving force behind the disparity.
Donadson and Shackley	1997	WTP	Payment scale	questionnaire	117 patients	77%	Used WTP as a means of testing empirically if 'process' utility exists when comparing laparoscopic versus conventional cholecystectomy. The results lead to the refection of the hypothesis that information on process of care would lead to higher WTP. The authors comment that due to the design of the study and the difficulties in defining process and outcome, it cannot be concluded that process utility does not exist.
Davey P et al.	1998	WTP	Open-ended and bidding game	interviews	83 patients	Not reported	Use of WTP to place an economic value on Insulin Lispro versus Neutral (regular) insulin therapy. Overall 92% of the sample preferred insulin lispro and 8% preferred neutral insulin. The incremental benefit per patient was calculated as 452.16 Australian dollars per year.
Olsen J and Donaldson C	1998	WTP	Payment scale	Interview	150 from general population	53%	Asked 150 interviewees their WTP in increased earmarked taxation for three health care programmes: a helicopter ambulance service, more heart operations and more hip replacements. Comparisons were made of WTP for these programmes and the health outcome in terms of QALYs.
Bala et al.	1998	WTP/ QALYs	Double bound closed ended	interviews	114 general population	16.5%	Compare QALYs amd WTP in eliciting individuals' preferences for health outcomes associated with shingles. Found no significant correlation between QALYs and WTP across individuals. The results raise questions about whether QALYs and WTP are equivalent preference-based measures of health outcomes.
Donaldson et al.	1998	WTP	Open-ended	questionnaire	113 patients	75%	Assess the feasibility of the use of WTP as a measure of the benefits of intrapartum care. The study compared intrapartum care in a midwife-managed delivery unit versus care in a consultant led labour ward. Most women expressed a preference for care in a midwives unit. However, strength of preference, as reflected by WTP, was greater among those in the smaller grp, who expressed a

							preference for care in a consultant-led ward.
Zethraeus N	1998	WTP	Closed ended	interview	104 women	Not reported	Aims to estimate the WTP for hormone replacement therapy (HRT) to alleviate menopausal symptoms. Uses TTO and RS to estimate the QALY gain from HRT. The WTP for HRT can then be used to calculate a WTP per QALY gained from HRT. The mean WTP is above the mean treatment cost for HRT.
Blumenschein and Johannesson	1998	WTP	Closed ended and bidding game	questionnaire	69 patients	100%	The aim was to quantitate the quality of life effects of asthma. The rating scale, time trade off, standard gamble and WTP techniques were adopted to measure the effects of an asthma cure. Nearly all correlations between the dimensions of quality of life, health state utilities, and WTP were in the expected direction.
Gibb et al.	1998	WTP	Payment scale	interview	50 patients	100%	This study measured women's strength of preference for medical abortion versus surgical vacuum aspiration. 64% preferred to have the medical method. The amounts offered for each method were similar; however, a minority gave higher values for the medical method. Validity of the technique was supported by the finding of a positive association with social class and the importance women attached to having choice.
Ortega A et al.	1998	WTP	Payment scale	Questionnaire and interview	100 patients 50 general population	94% (patients) 31% (general population)	Use of the WTP technique to estimate the net cost or benefit of administering prophylactic epoetin alfa to patients. Asked a cohort of 100 patients for their WTP value and also a sample for 50 people from the general population to get a societal WTP.
Lee SJ et al.	1998	WTP	Closed ended	questionnaire	412 patients	Not reported	Tested the impact of risk information on patients' WTP for autologous blood donations. Found that the WTP was sensitive to perceived transfusion risk, personal income, and dread of transfusions.
O'Brien B et al.	1998	WTP/WTA	Bidding Game	Interviews	220 enrollees	14% responded 6% interviewed	Use of the bidding game approach to assess the WTP for a new drug administered after chemotherapy treatment for cancer. Done from a user based and an insurance based perspective. No evidence was found of significant starting point bias. WTA exceeds WTP for coverage of the same

							benefit.
Dixon S and Shackley P	1998	WTP/WTA	Payment scale	Interview	100	100%	Two alternative measures were used to elicit public preferences for water fluoridation. Half the questionnaires contained 'WTA compensation' questions and the other half 'WTP to prevent' the loss.WTP to prevent questions were easier to answer than the WTA compensation questions. However the method performed no better. The authors do express a concern about the large no. of protests in the study.
Dranitsaris G	1999	WTP	Payment scale	interview	100 subjects	56%	Taking a societal perspective, 100 Canadian taxpayers were asked for their maximum WTP for the benefits of pamidronate (drug used to treat skeletal-related events associated with multiple myeloma (disorder of bone marrow)). Respondents stated that they would be WTP an average of Can $3,364 as an income tax increase to be paid over their lifetime. The WTP instrument was simple to administer and easily understood by the respondents.
Mathews DC et al.	1999	WTP	Bidding Game	Interviews	42 subjects	77.7%	A pilot study designed to test the feasibility and validity of using the WTP method in a dental setting. Preferences for alternative treatments for periodontal disease and the corresponding WTP values were elicited. It was found that periodontal surgery was the preferred treatment for moderate to advanced periodontal disease and the questionnaire had an accurate representation of treatment and outcomes, establishing face and content validity.
Barner et al.	1999	WTP	Bidding Game	Telephone interview	116 patients	29.7%	Assessment of asthma patients' WTP and willingness to give up time for an asthma self-management program. Mean patient WTP was $29.50 and patients reported that they were willing to spend a mean of 5.8 hours per week on the program. The study identified several factors that may affect WTP and willingness to spend time such as indicated interest in participation, greater perceived access to health care, previous participation in a program, among others.
Sorum PC	1999	WTP	Open-ended	questionnaire	322 parents of sick children	68%	Parents' preferences for the events and outcomes associated with acute otitis media (AOM) and its treatment were

							quantified by means of a questionnaire asking how much they would be WTP for them. The WTP values were then used to calculate the expected value (EV) of treating or not treating AOM with antibiotics. The EV of treating with antibiotics was, robustly superior because of parents' desire to avoid any increased risk of their children's death.
Lundberg L et al	1999	WTP	Closed-ended and Bidding Game	Questionnaire and Interview	366 patients	96%	Using data from 366 patients the study examined the correlation between willingness to pay and different dimensions of health-related quality of life, as measured by general and disease specific quality of life instrument and a subjective estimate of disease severity. Results showed that willingness to pay was correlated with the Dermatology quality of life index and disease activity but not with SF-36.
Larson RA	2000	WTP	Not specified	Mail questionnaire	175 subjects	7.6%	Use of WTP to measure patients demand for comprehensive pharmaceutical care services. Respondents were WTP a mean of $13 for a one-time consultation and $28 for this plus 1 year of monitoring. The study concludes that the majority of patients are WTP for pharmaceutical care even if they are not presently receiving that level of care.
Liu et al.	2000	WTP	Closed-ended with follow-up	Interview	650 mothers	92.8%	Use of CV method to estimate mothers WTP for her own and her child's health. Mothers of primary school students were interviewed using a closed-ended with follow up question design. The average mother is WTP more to protest her child than herself from suffering a cold. Median WTP to avoid the average mother's and child's colds are US$37 and US$57, respectively.
Dranitsaris G et al	2000	WTP	Open-ended	Interviews	125 subjects	64%	The study uses WTP to measure the economic value of a new insulin formation. WTP values were elicited for the new insulin drug versus an alternative therapy with approximately 84% of the sample preferring the new drug. The mean WTP for the new drug was $Can 35.28 which was at least 2-fold higher than the incremental monthly cost of the alternative drug.
Poyner TF, Menday AP,	2000	WTP	'checklist method'	questionnaire	258 patients	Not reported	Use of WTP to assess preferences for treatment of chronic plaque psoriasis (calcipotriol versus dithranol). Overall

Williams ZV							satisfaction with calcipotriol was significantly better however the difference in preference for calcipotriol over dithranol was not picked up by the difference in WTP values. Patients were WTP a mean of £12.16 monthly for calcipotriol and £10.66 monthly for dithranol. Overall WTP did not correlate well with overall treatment satisfaction nor with household income.
Narbro K and Sjostrom L	2000	WTP	Open-ended	questionnaire	3,558 patients	88.4%	To estimate WTP for effective treatment in a sample of obese individuals. Average WTP was SEK 26,900 and the median value was SEK 10,900. High WTP was associated with high income, high weight, high education, female sex, poor perceived health, low current age and low age at onset of obesity. The study concludes that most obese patients are WTP approximately twice their monthly salary for treatment of obesity.
Zarkin et al.	2000	WTP	Payment scale	Mail questionnaire	393 subjects	Not clear	Use of WTP to assess both the tangible and the intangible benefits of drug abuse treatment. WTP was estimated for two different types of drug treatment programs: a program for all drug users and a program specifically targeted to woman drug users. The mean WTP for both types of drug treatment programs was estimated to be approximately $37 per respondent.
Keith PK et al.	2000	WTP	Open ended	Questionnaire	241 (patients)	99%	The WTP method was adopted to conduct a CBA on intranasal budesonide for seasonal allergic rhinitis. The study assessed WTP before and then after the study. They found that there was a mean reduction of $2.86 in the WTP. They also found that there was no significant relationship between income and WTP.
Stephens M et al.	2000	WTP	Open ended	12-item survey?	137 (patients)	91%	To establish womens' WTP for a sonogram during pregnancy. The study took place in a naval medical centre and all women were eligible for care. Results showed that 98% of the sample wanted a sonogram. 51% were willing to go elsewhere to get a sonogram and would be WTP $85.
Slothuus U and Brooks RG	2000	WTP	Contingent ranking and	Interview	120 (outpatients)	67%	Subjects were asked for their WTP for alleviation of symptoms that may be the outcome of a drug

			closed ended.				(antirheumatic agent: cA2) used to treat athiritis. The study compared two approaches of asking WTP: contingent ranking and closed ended. There was no statistical difference in the results between the two methods.
Kupperman M et al.	2000	WTP	Open ended	Interview	206 (families of patients)	64%	Using WTP, investigate parents' preferences for various outcomes associated with childhood vaccinations. Study found that the WTP to avoid less desired outcomes increased.
Hirth R et al	2000	WTP	Payment scale	Questionnaire	53 patients 34 physicians 26 executives	100%	Surveyed patients, physicians and managed care executives to assess their WTP for diagnostic certainty for peptic ulcer disease and gastroesophageal reflux disease. Participants were also asked to estimate the WTP of each of the other types of respondents. Patients were most likely and executives least likely to value diagnostic certainty.
Bishai DM and Hui Chu Lang	2000	WTP	Closed ended	Telephone interview	159 patients	70%	Using closed ended approach the authors estimate the demand curves for a one month reduction in waiting time for cataract surgery based on survey data collected in Manitoba, Barcelona and Denmark. Barcelonan patients have greater WTP for shortened waiting time than the Danes and Manitobans.
Eastaugh S R	2000	WTP	Bidding game	Mail survey	223 patients	77%	Assessed the WTP for treatment of bleeding disorders. Randomly assigned half of the sample to receive a low initial bid and the other half to receive a high bid. Those who received a low bid gave a lower overall WTP compared to those who received a high bid. WTP was influenced by income, education and a category rating scale for health status.
Taylor SJ and Armour CL	2000	WTP	Payment scale	questionnaire	359	89%	Measurement of consumer preference for treatments used to induce labour. 73.7% of patients preferred gel. The mean maximum WTP for amniotomy plus oxytocin was Aus$133 while that for gel was Aus$178. Using the WTP method consumers were able to assess drug information, make an informed choice and to value that choice.
Cho-Min-Naing et al	2000	WTP	Bidding Game	interview	1000	Not reported	Assessed the WTP *ex post* and *ex ante* for the ICT Malaria *Pf/Pv* test kit in two villages in Myanmar. WTP was higher

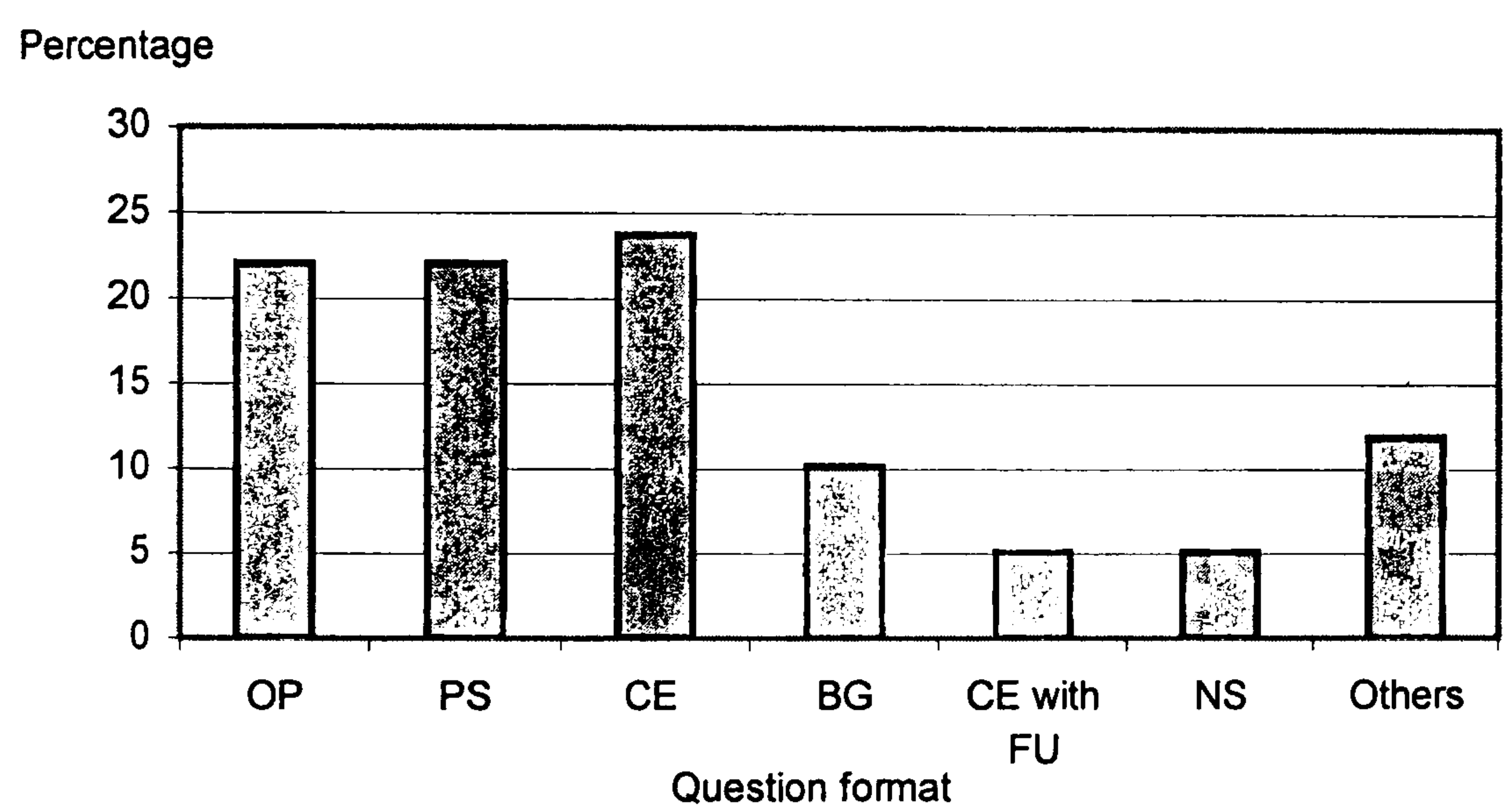
							in *ex post* conditions than in *ex ante* conditions ($p<0.001$). WTP was significantly positively associated with the average monthly income of the respondents and severity of illness in both *ex post* and *ex ante* conditions ($p<0.001$).
Ryan M, San Miguel F	2000	WTP	Payment scale	Questionnaire	75	51%	The paper sets out to develop a test for consistency in WTP experiments. The test is applied to elicit women's preferences for two alternative treatments of menorrhagia. Thirty percent of respondents fail the consistency test and cost-based responses were found to partly explain the inconsistent responses.
Slothuus M, Larsen M and Peter J	2000	WTP	CE/ CE with follow-up	Interviews	120 patients	67%	Comparison of closed-ended with closed-ended with follow-up applied to arthritis symptom alleviation. Tests show that the two different type of questions yield significantly different WTP amounts (DKK 637 v's DKK 1,268). Including a follow-up question increases the precision of the result. The authors recommend that when choosing between the two question designs, power, efficiency and size could be used as selection criteria.
Papatheofanis FJ	2000	WTP	Closed ended	Questionnaire	87 patients	27%	Willingness to pay for positron emission tomography imaging by patients with suspected benign or malignant lung disease. Patients with increased perception of risk were WTP more than those with lower perceived risk. Study concludes that individuals are WTP additional out of pocket costs for diagnostic imaging to reduce their perception of risk and improve QOL.
Onwujekwe O	2001	WTP	Bidding game and closed-ended with follow-up	interview	800 households	95%	Aim of study was to compare the theoretical and predictive validity of the bidding game versus the closed ended with follow-up elicitation format. The study was set in Nigeria and elicited WTP for insecticide treated nets. The bidding game format produced high mean and median WTP amounts compared to the closed ended with follow up. The paper concludes that the main result of the study is a movement towards developing an elicitation technique that fits in with the study setting (i.e. developing country, Nigeria).

Blumenschein et al.	2001	WTP	Closed-ended	interview	173 patients (84 hypothetical grp, 89 in real grp)	Not reported	Study designed to assess the difference between hypothetical and real purchase decisions for a pharmacist provided asthma management program. Out of the hypothetical grp, 38% of patients said that they would purchase the programme compared to 12% in the real group. The study concludes that the closed-ended format does lead to an overestimation of WTP values but this can be modified by distinguishing the ‘definitely sure’ responses.
Smith RD	2001	WTP	Open-ended	Interview	47 general population	94%	Investigation into the discriminant validity of TTO technique versus the WTP technique. Use of the open-ended WTP format with a prompt after 60 seconds. Found WTP to be more sensitive between different dimensions of health at the same nominal level of health status and between different levels of health within each dimension. The author tentatively suggests that WTP seems to be a more sensitive measure of change in health status.

It is evident that the most popular direction of measurement applied is the WTP over the WTA estimate. Table A1.1 describes 64 studies that have used the CV method, only three of them have attempted to estimate the WTA value.

The choice of elicitation format varies across the studies presented in the table. Figure A1.1 displays the distribution of question formats across the studies.

Figure A1.1 Popularity of question format



The closed-ended method marginally emerges as the most popular method however both the open-ended and payment scale formats have commonly been used. Given that the choice of format is influenced by the study design (difficult to administer a bidding game design using mail questionnaires) it is interesting to compare the question format by the instrumentation method. Table A1.2 illustrates the results.

Table A1.2 Question format by study design

Question format	Choice of instrumentation format		
	Questionnaire (%)	Interview (%)	Others (%)
OP	35	16	2
PS	27	16	25
CE	19	28	25
BG	4	16	13
CEFU	0	8	11
NS	4	8	
Others	12	8	25
Total	100	100	100

Within the studies that used a mail questionnaire as the study design, 35% adopted the open-ended question followed by 27% who adopted the payment scale. This is to be expected given the ease of use with these question designs within a questionnaire format compared to the closed-ended and bidding game approaches. As expected, the closed-ended becomes the most popular technique amongst those studies that used personal interviews to elicit WTP whilst the open-ended, payment scale and bidding game are equally adopted.

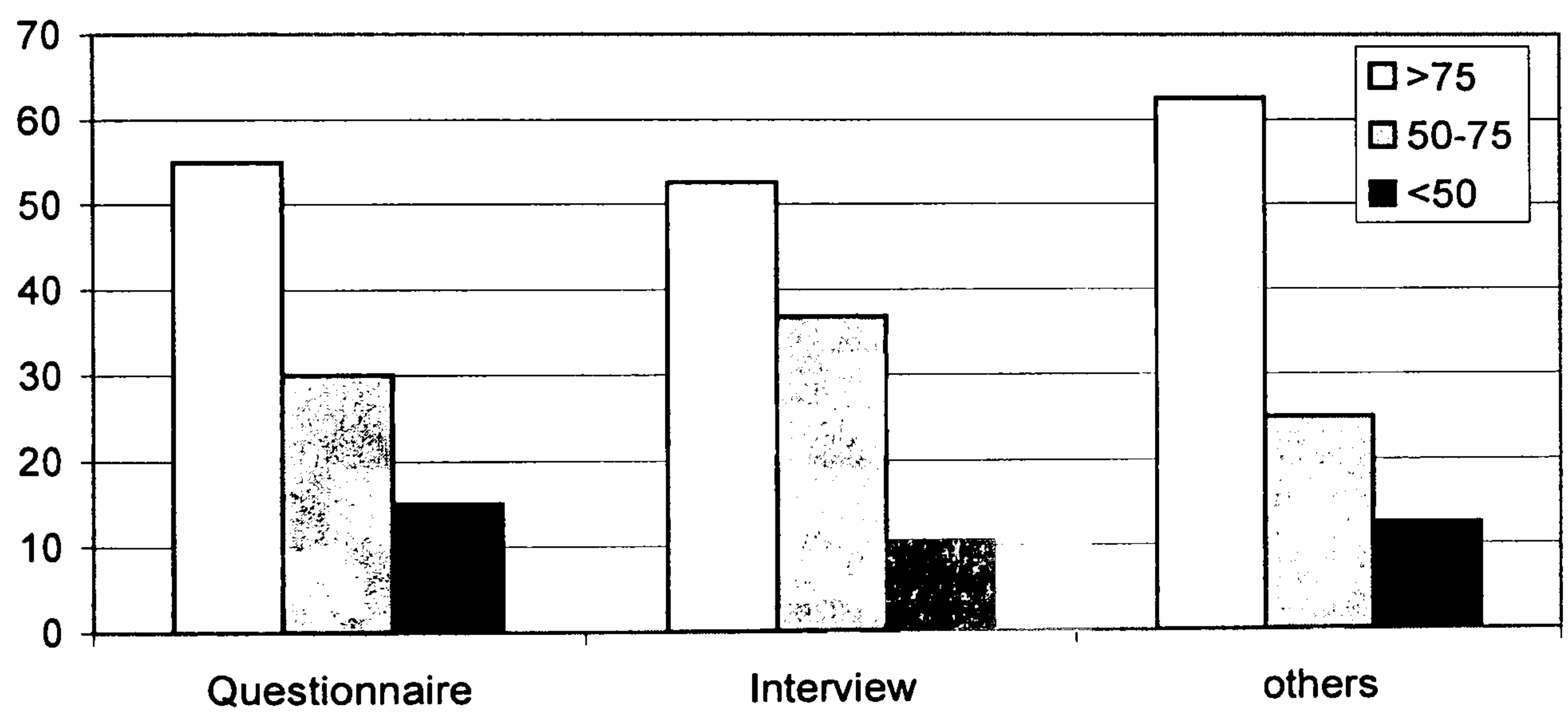
As table A1.3 shows, none of the studies that adopted particular question formats had a larger than average sample size. It seems therefore that predicted sample size does not influence choice of question format.

Table A1.3 Question format by sample size

Question format	Sample size		Total
	<200 (%)	>=200 (%)	
OP	21	24	22
PS	29	12	22
CE	26	20	23
BG	9	12	10
CEFU	3	8	5
NS	6	4	5
Others	6	20	13
Total	100	100	100

The most popular form of instrumentation technique across all the studies presented in table A1.1 are interviews (31) and mail questionnaires (31), only three studies used telephone interviews. Two studies did not specify how the WTP values were elicited. It is interesting to examine the impact the instrumentation format has on the response rate, figure A1.2 presents the results. It seems that the mail questionnaire and the personal interview have produced very similar results in terms of response rate from the participants. Table A1.4 presents the actual difference in sample size by instrumentation format.

Figure A1.2 Response rate by instrumentation format (%)



As you would expect the majority of studies that adopted the use of mail questionnaires achieved a sample size of >200, this is in contrast to those studies that adopted personal interviews as the instrumentation format as only 28% of them achieved a sample size of >200.

Table A1.4 Sample size by instrumentation format

	Sample size		
	<=200	>200	Total
Questionnaires (%)	46	54	100
Personal interviews (%)	72	28	100
Others (%)	58	42	100

As discussed, the contingent valuation method is a valuable technique to measure the strength of preference for treatment or prevention of a disease. This is apparent in table A1.1 as there are 11 studies that have applied the technique to estimate WTP for two alternative treatments, i.e. WTP for inpatient versus day-case surgery or treatments for chronic plaque psoriasis. What has emerged from describing the studies in table A1.1 are the disease areas in which the CV technique has been applied; paediatric care, arthritis, asthma programmes, pharmaceutical care and ante and post-natal care along with using it to compare two alternative therapies for treatment of the same condition.

Appendix Two

The National Screening Committee (NSC) Criteria

The Criteria for appraising the viability, effectiveness and appropriateness of a screening programme

The criteria, which are set out below, are based on the classic criteria first promulgated in a WHO Report in 1966 but take into account both the more rigorous standards of evidence required to improve effectiveness and the greater concern about the adverse effects of healthcare; regrettably some people who undergo screening will suffer adverse effects without receiving benefit from the programme.

These criteria have been prepared taking into account international work on the appraisal of screening programmes, particularly that in Canada (2) and the United States (3). It is recognised that not all of the Criteria and questions raised in the Format will be applicable to every proposed programme, but as many as possible should be answered since this will assist the NSC to make quicker and better evidence based decisions.

All of the following criteria should be met before screening for a condition is initiated:

The condition

1.1. The condition should be an important health problem.
1.2. The epidemiology and natural history of the condition, including development from latent to declared disease, should be adequately understood and there should be a detectable risk factor, or disease marker and a latent period or early symptomatic stage.
1.3. All the cost-effective primary prevention interventions should have been implemented as far as practicable.

The test

1.4. There should be a simple, safe, precise and validated screening test.
1.5. The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.
1.6. The test should be acceptable to the population.
1.7. There should be an agreed policy on the further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals.

The treatment

1.8. There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.
1.9. There should be agreed evidence based policies covering which individuals

should be offered treatment and the appropriate treatment to be offered.
1.10. Clinical management of the condition and patient outcomes should be optimised by all health care providers prior to participation in a screening programme.

The screening programme

1.11. There must be evidence from high quality Randomised Controlled Trials that the screening programme is effective in reducing mortality or morbidity.
Where screening is aimed solely at providing information to allow the person being screened to make an "informed choice" (e.g. Down's syndrome, cystic fibrosis carrier screening), there must be evidence from high quality trials that the test accurately measures risk. The information that is provided about the test and its outcome must be of value and readily understood by the individual being screened.
1.12. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public.
1.13. The benefit from the screening programme should outweigh the physical and psychological harm (caused by the test, diagnostic procedures and treatment).
1.14. The opportunity cost of the screening programme (including testing, diagnosis, treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (i.e. value for money).
1.15. There must be a plan for managing and monitoring the screening programme and an agreed set of quality assurance standards.
1.16. Adequate staffing and facilities for testing, diagnosis, treatment and programme management should be made available prior to the commencement of the screening programme.
1.17. All other options for managing the condition should have been considered (e.g. improving treatment, providing other services), to ensure that no more cost effective intervention could be introduced or current interventions increased within the resources available.
1.18. Evidence-based information, explaining the consequences of testing, investigation and treatment, should be made available to potential participants to assist them in making an informed choice.
1.19. Public pressure for widening the eligibility criteria for reducing the screening interval, and for increasing the sensitivity of the testing process, should be anticipated. Decisions about these parameters should be scientifically justifiable to the public.

References:

Department of Health. *Screening of pregnant women for hepatitis B and immunisation of babies at risk.* London: Department of Health, 1998. (Health Service Circular: HSC 1998/127)
Wilson JMG, Jungner G. Principles and practice of screening for disease. Public Health Paper Number 34. Geneva: WHO, 1968.
Cochrane AL. Holland WW. Validation of screening procedures. Br Med Bull. 1971,

27, 3.
Sackett DL, Holland WW. Controversy in the detection of disease. Lancet 1975; 2:357-9.
Wald NJ (Editor). Antenatal and Neonatal screening. Oxford University Press, 1984.
Holland WW, Stewart S. Screening in Healthcare. The Nuffield Provincial Hospitals Trust, 1990.
Gray JAM. Dimensions and definitions of screening. Milton Keynes: NHS Executive Anglia and Oxford, Research and Development Directorate, 1996.

Appendix Three

YOUR OPINION ON HEALTH SCREENING

Screening For Colorectal Cancer

Cancer of the colon or rectum - commonly referred to together as colorectal or bowel-cancer is the second most common cancer in the UK. It represents one of the most significant causes of cancer deaths in industrialised countries.

However, there are now grounds for believing that a national screening programme aimed at detecting colorectal cancer at a very early stage could save lives. There are several different screening techniques available for detecting colorectal cancer, these are currently being assessed in the UK.

We are trying to get an idea about how people would value a national screening programme for colorectal cancer and would appreciate your opinion. This project is being sponsored by the Medical Research Council and your answers will be used for research purposes only.

In order to help us, we would be grateful if you could take the time to complete the enclosed questionnaire. You do not have to do so, and your care will not be affected in any way if you decide not to take part in this survey.

There are no right or wrong answers to the questions asked. We are interested in your view. Please take time to read the descriptions of the disease and the screening investigations before attempting to answer the questions.

Some questions will ask you to tick a box like this ☑. Other questions will ask you to write your answer in the space provided. Please use black or blue ink.

This questionnaire is anonymous.
Your answers will be treated as confidential.
If this questionnaire causes concern, please contact Jane Wolstenholme or Emma Frew at the address below, or your GP. Complaints about the way the study is being conducted should be addressed to your Local Authority Complaints Department, (address{please see attached sheet}).**

Contact name & address:
Jane Wolstenholme or Emma Frew
Trent Institute for Health Services Research
Room B39, Medical School,
QMC,
Nottingham. NG7 2UH

tel: 0115 9709765

A Questionnaire to Assess the Value to You of Screening for Colorectal Cancer.

Colorectal Cancer:
Cancer of the colon or rectum - commonly referred to together as colorectal or bowel cancer is the second commonest cancer in the UK.

Why screen for colorectal cancer?
Screening means doing simple tests to pick up hidden problems. If people wait for symptoms before seeking care, there will be a high proportion of more advanced cancers detected, which are more difficult to treat. Patients whose cancers are detected at earlier stages by screening do much better than patients with more advanced cancers.

What does screening involve?
There are two main types of screening test for bowel cancer, neither are available in the UK at present. Trials are being carried out to assess whether it would be worth introducing a national screening programme for colorectal cancer using one of the screening tests outlined below.

A) <u>Faecal Occult Blood (FOB) test</u>
The faecal occult blood test would involve screening <u>once every two years</u> from the age of 50 to 74 years old. The test is carried out by yourself at home and involves a small kit which would be sent to you in the post. It requires small stool (faeces) samples to be taken over a time period of three days, placed on special paper within the test kit and then returned to your local GP. Investigation of the test kit would enable the confirmation of the possible presence or absence of bowel cancer. If the test confirms the possible presence of cancer, you would be required to come into your local hospital for further investigations.

B) <u>Flexi-scope test</u>
The Flexi-scope screening test would involve a <u>one-off</u> screen for all people aged 60-65 years old. The flexi-scope is a thin flexible tube with a tiny camera on the end. A doctor inserts the Flexi-scope into the back passage and looks for bowel polyps. Most bowel polyps are harmless growths, but some of them may become cancerous. The majority of polyps can be removed quickly and painlessly with the Flexi-scope. Some of the polyps may need to be removed using a colonoscope under general anaesthetic. Removing these polyps helps to prevent bowel cancer. The Flexi-scope would be done at a local hospital clinic, no anaesthetic is required. An enema would be provided for you to use at home before the test to clear your bowel. The test would only take 5 minutes although it would be necessary to devote an hour to the visit because of waiting time. If an abnormality is found this can be often removed using the Flexi-scope during the same visit.

1) Would you have one of these screening tests for bowel cancer if it was available?
(please tick the appropriate box)

Yes	No	Don't know
❑	❑	❑

IF YOU TICKED 'YES' OR 'DON'T KNOW', PLEASE GO TO **QUESTION 2**

IF YOU TICKED 'NO', PLEASE GO TO **QUESTION 6**

2) Which type of method of screening would you <u>prefer</u> if it was available to you?
*(please tick <u>**one**</u> appropriate box)*

Faecal Occult Blood Test	Flexi-scope	No preference
❑	❑	❑

One way of measuring the value of screening for colorectal cancer is to ask you what you would be prepared to give up to receive this service i.e. how much money you would be willing to pay for it. Of course, if colorectal cancer screening did come into existence in England, the screening test would be provided <u>free</u> by the NHS. <u>**We also believe that people should not have to pay for health care This is simply a method of measuring how strongly you feel about having a new screening programme and how much you would value such a service.**</u>

There are no right or wrong answers. The amount you say could be large or small. It is up to you! We are interested in your view.

3) What is the maximum amount of money you would be willing to pay for having the <u>complete series of FOB tests every 2 years from the age of 50 until the age of 74?</u>

(please write your answer in the space below)

£ ____________________

4) What is the maximum amount of money you would be willing to pay for the <u>one-off Flexi-scope test at the age of 65?</u>

(please write your answer in the space below)

£ ____________________

5) **Could you please explain the reasons for your answers to questions 3 and 4?**

(please write your explanation in the space below)

6) We would now like to ask you some questions about yourself. To repeat, these details are entirely confidential.

Gender
(please tick the appropriate box)

❑ Female
❑ Male

Could you please tell us your age?
(please write your answer in the box) ☐ **years**

Ethnic origin:
Please tick the box against the term which describes you:

❑ White
❑ Black - African
❑ Black - Caribbean
❑ Black - Other
❑ Indian
❑ Pakistani
❑ Bangladeshi
❑ Chinese
❑ Asian - Other
❑ Other (please specify) ....

What is your marital status?
(please tick the appropriate box)

❑ single
❑ married/living with partner
❑ widowed

How many children do you have?
(please write your answer in the box) ☐

Are you currently:
(please tick the appropriate box)

❑ Employed
❑ Unemployed
❑ House-wife/house-husband
❑ Retired
❑ Student
❑ Other
If other please state: ....

Current or most recent occupation: ....

Are you the main income earner in the household?
(please tick the appropriate box)

❑ Yes
❑ No

If **NO**, what is the occupation of the main income earner?...

Could you please estimate the annual income of your household before deducting tax and national insurance? (If you receive any benefits or pensions include them as income).

(please tick the appropriate box)

- ☐ Less than £10,000
- ☐ £10,000 - £20,000
- ☐ £20,001 - £30,000
- ☐ More than £30,000

What was your age when you left full-time education?

(please write your answer in the box)

☐ **years**

Have you or your close family ever suffered from any of the following health problems?

(please tick the appropriate box)

Yourself	Your family	
☐	☐	Stomach problems e.g. ulcer
☐	☐	Piles or haemorrhoids
☐	☐	Heart trouble
☐	☐	Cancer
☐	☐	Stroke
☐	☐	Severe depression

Would you say your health is:

(please tick the appropriate box)

- ☐ Excellent?
- ☐ Good?
- ☐ Fair?
- ☐ Poor?

Regarding smoking, would you describe yourself as :

(please tick the appropriate box)

- ☐ A smoker
- ☐ An ex-smoker
- ☐ Never smoked

How many times have you visited your GP in the last year?
(please tick the appropriate box)

- ❑ None
- ❑ Once
- ❑ Twice
- ❑ Three
- ❑ More than three times

How many times have you visited your dentist in the past two years?
(please tick the appropriate box)

- ❑ None
- ❑ Once
- ❑ Twice
- ❑ Three times
- ❑ Four times
- ❑ More than four times

What other diseases have you been screened for in the last 5 years?
(please tick the appropriate box)

- ❑ None
- ❑ Cervical cancer
- ❑ Breast cancer
- ❑ Other

  If other, please state: ....

In general, how important do you feel the following health measures are?
(please tick the appropriate box)

	Not at all	Somewhat	Moderately	Very	Extremely
to exercise regularly	❑	❑	❑	❑	❑
to eat plenty of fruit	❑	❑	❑	❑	❑
for a women to have a cervical smear at least once every 5 years	❑	❑	❑	❑	❑
for a women to have a breast screen every 3 years	❑	❑	❑	❑	❑

How worried are you about getting bowel cancer?
(please tick the appropriate box)

Not worried	A bit worried	Quite worried	Very worried
❑	❑	❑	❑

Compared with other men and women of your age, do you think your chances of getting bowel cancer are:
(please tick the appropriate box)

Much lower	A little lower	About the same	A little higher	Much higher
❑	❑	❑	❑	❑

Although screening for colorectal cancer is not routinely available in the UK, some people may already have already received an FOB or Flexi-scope test. If you have please indicate below.

❑ FOB
❑ Flexi-scope

WE WOULD LIKE TO TAKE THIS OPPORTUNITY TO REMIND YOU THAT YOUR ANSWERS ARE CONFIDENTIAL AND THAT THEY WILL NOT BE USED IN SETTING CHARGES FOR SUCH SERVICES

PLEASE PLACE THE QUESTIONNAIRE IN THE ENVELOPE PROVIDED

MANY THANKS FOR YOUR VALUABLE CONTRIBUTION TO OUR STUDY

Appendix Four

YOUR OPINION ON HEALTH SCREENING

Screening For Colorectal Cancer

Cancer of the colon or rectum - commonly referred to together as colorectal or bowel-cancer is the second most common cancer in the UK. It represents one of the most significant causes of cancer deaths in industrialised countries.

However, there are now grounds for believing that a national screening programme aimed at detecting colorectal cancer at a very early stage could save lives. There are several different screening techniques available for detecting colorectal cancer, these are currently being assessed in the UK.

We are trying to get an idea about how people would value a national screening programme for colorectal cancer and would appreciate your opinion. This project is being sponsored by the Medical Research Council and your answers will be used for research purposes only.

In order to help us, we would be grateful if you could take the time to complete the enclosed questionnaire. You do not have to do so, and your care will not be affected in any way if you decide not to take part in this survey.

There are no right or wrong answers to the questions asked. We are interested in your view. Please take time to read the descriptions of the disease and the screening investigations before attempting to answer the questions.

Some questions will ask you to tick a box like this ☑. Other questions will ask you to write your answer in the space provided. Please use black or blue ink.

This questionnaire is anonymous.
Your answers will be treated as confidential.
If this questionnaire causes concern, please contact Jane Wolstenholme or Emma Frew at the address below, or your GP. Complaints about the way the study is being conducted should be addressed to your Local Authority Complaints Department, (address{please see attached sheet}).**

Contact name & address:
Jane Wolstenholme or Emma Frew
Trent Institute for Health Services Research
Room B39, Medical School,
QMC,
Nottingham. NG7 2UH

tel: 0115 9709765

A Questionnaire to Assess the Value to You of Screening for Colorectal Cancer.

Colorectal Cancer:

Cancer of the colon or rectum - commonly referred to together as colorectal or bowel cancer is the second commonest cancer in the UK.

Why screen for colorectal cancer?

Screening means doing simple tests to pick up hidden problems. If people wait for symptoms before seeking care, there will be a high proportion of more advanced cancers detected, which are more difficult to treat. Patients whose cancers are detected at earlier stages by screening do much better than patients with more advanced cancers.

What does screening involve?

There are two main types of screening test for bowel cancer, neither are available in the UK at present. Trials are being carried out to assess whether it would be worth introducing a national screening programme for colorectal cancer using one of the screening tests outlined below.

A) <u>Faecal Occult Blood (FOB) test</u>

The faecal occult blood test would involve screening <u>once every two years</u> from the age of 50 to 74 years old. The test is carried out by yourself at home and involves a small kit which would be sent to you in the post. It requires small stool (faeces) samples to be taken over a time period of three days, placed on special paper within the test kit and then returned to your local GP. Investigation of the test kit would enable the confirmation of the possible presence or absence of bowel cancer. If the test confirms the possible presence of cancer, you would be required to come into your local hospital for further investigations.

B) <u>Flexi-scope test</u>

The Flexi-scope screening test would involve a <u>one-off</u> screen for all people aged 60-65 years old. The flexi-scope is a thin flexible tube with a tiny camera on the end. A doctor inserts the Flexi-scope into the back passage and looks for bowel polyps. Most bowel polyps are harmless growths, but some of them may become cancerous. The majority of polyps can be removed quickly and painlessly with the Flexi-scope. Some of the polyps may need to be removed using a colonoscope under general anaesthetic. Removing these polyps helps to prevent bowel cancer. The Flexi-scope would be done at a local hospital clinic, no anaesthetic is required. An enema would be provided for you to use at home before the test to clear your bowel. The test would only take 5 minutes although it would be necessary to devote an hour to the visit because of waiting time. If an abnormality is found this can be often removed using the Flexi-scope during the same visit.

1) Would you have one of these screening tests for bowel cancer if it was available?
(please tick the appropriate box)

Yes	No	Don't know
❑	❑	❑

IF YOU TICKED 'YES' OR 'DON'T KNOW', PLEASE GO TO **QUESTION 2**

IF YOU TICKED 'NO', PLEASE GO TO **QUESTION 6**

2) Which type of method of screening would you <u>prefer</u> if it was available to you?
*(please tick <u>**one**</u> appropriate box)*

Faecal Occult Blood Test	Flexi-scope	No preference
❑	❑	❑

One way of measuring the value of screening for colorectal cancer is to ask you what you would be prepared to give up to receive this service i.e. how much money you would be willing to pay for it. Of course, if colorectal cancer screening did come into existence in England, the screening test would be provided <u>**free**</u> by the NHS. <u>**We also believe that people should not have to pay for health care This is simply a method of measuring how strongly you feel about having a new screening programme and how much you would value such a service.**</u>

There are no right or wrong answers. The amount you say could be large or small. It is up to you! We are interested in your view.

3) What is the maximum amount of money you would be willing to pay for having the complete series of FOB tests every 2 years from the age of 50 until the age of 74?

Put a ✓ next to the amounts that you are sure you **would** pay

Put a **X** next to the amounts that you are sure you **would not** pay

Put a circle around the **maximum amount** you would pay

£0
£5
£10
£20
£30
£40
£50
£60
£70
£80
£90
£100
£120
£140
£160
£180
£200
£250
£300
£350
£400
£450
£500
£600
£700
£800
£900
£1000 +

(If £1000 + state the exact amount in the space provided below)

£ ____________________

4) What is the maximum amount of money you would be willing to pay for the one-off Flexi-scope test at the age of 65?

Put a ✓ next to the amounts that you are sure you **would** pay

Put a **X** next to the amounts that you are sure you **would not** pay

Put a circle around the **maximum amount** you would pay

£0
£5
£10
£20
£30
£40
£50
£60
£70
£80
£90
£100
£120
£140
£160
£180
£200
£250
£300
£350
£400
£450
£500
£600
£700
£800
£900
£1000 +

(If £1000 + state the exact amount in the space provided below)

£ ____________________

5) Could you please explain the reasons for your answers to questions 3 and 4?

(please write your explanation in the space below)

6) We would now like to ask you some questions about yourself. To repeat, these details are entirely confidential.

Gender
(please tick the appropriate box)
❑ Female
❑ Male

Could you please tell us your age? ☐ **years**
(please write your answer in the box)

Ethnic origin:
Please tick the box against the term which describes you:

❑ White	❑ Pakistani
❑ Black - African	❑ Bangladeshi
❑ Black - Caribbean	❑ Chinese
❑ Black - Other	❑ Asian - Other
❑ Indian	❑ Other (please specify) ....

What is your marital status?
(please tick the appropriate box)

single	married/living with partner	widowed
❑	❑	❑

How many children do you have? ☐
(please write your answer in the box)

Are you currently:
(please tick the appropriate box)
❑ Employed
❑ Unemployed
❑ House-wife/house-husband
❑ Retired
❑ Student
❑ Other
If other please state: ....

Current or most recent occupation: ....

Are you the main income earner in the household?
(please tick the appropriate box)
❑ Yes
❑ No

If **NO**, what is the occupation of the main income earner?..

Could you please estimate the annual income of your household before deducting tax and national insurance? (If you receive any benefits or pensions include them as income).
(please tick the appropriate box)

Less than £10,000	£10,000 - £20,000	£20,001 - £30,000	More than £30,000
❑	❑	❑	❑

What was your age when you left full-time education?
(please write your answer in the box) ☐ **years**

Have you or your close family ever suffered from any of the following health problems?
(please tick the appropriate box)

Yourself	Your family	
☐	☐	Stomach problems e.g. ulcer
☐	☐	Piles or haemorrhoids
☐	☐	Heart trouble
☐	☐	Cancer
☐	☐	Stroke
☐	☐	Severe depression

Would you say your health is:
(please tick the appropriate box)

Excellent?	Good?	Fair?	Poor?
☐	☐	☐	☐

Regarding smoking, would you describe yourself as :
(please tick the appropriate box)

A smoker	An ex-smoker	Never smoked
☐	☐	☐

How many times have you visited your GP in the last year?
(please tick the appropriate box)

None	Once	Twice	Three	More than three times
☐	☐	☐	☐	☐

How many times have you visited your dentist in the past two years?
(please tick the appropriate box)

None	Once	Twice	Three times	Four times	More than four times
☐	☐	☐	☐	☐	☐

What other diseases have you been screened for in the last 5 years?
(please tick the appropriate box)

☐ None
☐ Cervical cancer
☐ Breast cancer
☐ Other
If other, please state: ....

In general, how important do you feel the following health measures are?
(please tick the appropriate box)

	Not at all	Somewhat	Moderately	Very	Extremely
to exercise regularly	❑	❑	❑	❑	❑
to eat plenty of fruit	❑	❑	❑	❑	❑
for a women to have a cervical smear at least once every 5 years	❑	❑	❑	❑	❑
for a women to have a breast screen every 3 years	❑	❑	❑	❑	❑

How worried are you about getting bowel cancer?
(please tick the appropriate box)

Not worried	A bit worried	Quite worried	Very worried
❑	❑	❑	❑

Compared with other men and women of your age, do you think your chances of getting bowel cancer are:
(please tick the appropriate box)

Much lower	A little lower	About the same	A little higher	Much higher
❑	❑	❑	❑	❑

Although screening for colorectal cancer is not routinely available in the UK, some people may already have already received an FOB or Flexi-scope test. If you have please indicate below.

❑ FOB
❑ Flexi-scope

WE WOULD LIKE TO TAKE THIS OPPORTUNITY TO REMIND YOU THAT YOUR ANSWERS ARE CONFIDENTIAL AND THAT THEY WILL NOT BE USED IN SETTING CHARGES FOR SUCH SERVICES

PLEASE PLACE THE QUESTIONNAIRE IN THE ENVELOPE PROVIDED

MANY THANKS FOR YOUR VALUABLE CONTRIBUTION TO OUR STUDY

Appendix Five

YOUR OPINION ON HEALTH SCREENING

Screening For Colorectal Cancer

Cancer of the colon or rectum - commonly referred to together as colorectal or bowel-cancer is the second most common cancer in the UK. It represents one of the most significant causes of cancer deaths in industrialised countries.

However, there are now grounds for believing that a national screening programme aimed at detecting colorectal cancer at a very early stage could save lives. There are several different screening techniques available for detecting colorectal cancer, these are currently being assessed in the UK.

We are trying to get an idea about how people would value a national screening programme for colorectal cancer and would appreciate your opinion. This project is being sponsored by the Medical Research Council and your answers will be used for research purposes only.

In order to help us, we would be grateful if you could take the time to complete the enclosed questionnaire. You do not have to do so, and your care will not be affected in any way if you decide not to take part in this survey.

There are no right or wrong answers to the questions asked. We are interested in your view. Please take time to read the descriptions of the disease and the screening investigations before attempting to answer the questions.

Some questions will ask you to tick a box like this ☑. Other questions will ask you to write your answer in the space provided. Please use black or blue ink.

This questionnaire is anonymous.
Your answers will be treated as confidential.
If this questionnaire causes concern, please contact Jane Wolstenholme or Emma Frew at the address below, or your GP. Complaints about the way the study is being conducted should be addressed to your Local Authority Complaints Department, (address{please see attached sheet}).**

Contact name & address:
Jane Wolstenholme or Emma Frew
Trent Institute for Health Services Research
Room B39, Medical School,
QMC,
Nottingham. NG7 2UH

tel:0115 9709765

A Questionnaire to Assess the Value to You of Screening for Colorectal Cancer.

Colorectal Cancer:
Cancer of the colon or rectum - commonly referred to together as colorectal or bowel cancer is the second commonest cancer in the UK.

Why screen for colorectal cancer?
Screening means doing simple tests to pick up hidden problems. If people wait for symptoms before seeking care, there will be a high proportion of more advanced cancers detected, which are more difficult to treat. Patients whose cancers are detected at earlier stages by screening do much better than patients with more advanced cancers.

What does screening involve?
There are two main types of screening test for bowel cancer, neither are available in the UK at present. Trials are being carried out to assess whether it would be worth introducing a national screening programme for colorectal cancer using one of the screening tests outlined below.

A) <u>Faecal Occult Blood (FOB) test</u>
The faecal occult blood test would involve screening <u>once every two years</u> from the age of 50 to 74 years old. The test is carried out by yourself at home and involves a small kit which would be sent to you in the post. It requires small stool (faeces) samples to be taken over a time period of three days, placed on special paper within the test kit and then returned to your local GP. Investigation of the test kit would enable the confirmation of the possible presence or absence of bowel cancer. If the test confirms the possible presence of cancer, you would be required to come into your local hospital for further investigations.

B) <u>Flexi-scope test</u>
The Flexi-scope screening test would involve a <u>one-off</u> screen for all people aged 60-65 years old. The flexi-scope is a thin flexible tube with a tiny camera on the end. A doctor inserts the Flexi-scope into the back passage and looks for bowel polyps. Most bowel polyps are harmless growths, but some of them may become cancerous. The majority of polyps can be removed quickly and painlessly with the Flexi-scope. Some of the polyps may need to be removed using a colonoscope under general anaesthetic. Removing these polyps helps to prevent bowel cancer. The Flexi-scope would be done at a local hospital clinic, no anaesthetic is required. An enema would be provided for you to use at home before the test to clear your bowel. The test would only take 5 minutes although it would be necessary to devote an hour to the visit because of waiting time. If an abnormality is found this can be often removed using the Flexi-scope during the same visit.

1) Would you have one of these screening tests for bowel cancer if it was available?
(please tick the appropriate box)

Yes	No	Don't know
❑	❑	❑

IF YOU TICKED 'YES' OR 'DON'T KNOW', PLEASE GO TO **QUESTION 2**

IF YOU TICKED 'NO', PLEASE GO TO **QUESTION 6**

2) Which type of method of screening would you <u>prefer</u> if it was available to you?
(please tick <u>one</u> appropriate box)

Faecal Occult Blood Test	Flexi-scope	No preference
❑	❑	❑

One way of measuring the value of screening for colorectal cancer is to ask you what you would be prepared to give up to receive this service i.e. how much money you would be willing to pay for it. Of course, if colorectal cancer screening did come into existence in England, the screening test would be provided **<u>free</u>** by the NHS. **<u>We also believe that people should not have to pay for health care This is simply a method of measuring how strongly you feel about having a new screening programme and how much you would value such a service.</u>**

There are no right or wrong answers. The amount you say could be large or small. It is up to you! We are interested in your view.

3) **What is the maximum amount of money you would be willing to pay for having the <u>complete series of FOB tests every 2 years from the age of 50 until the age of 74?</u>**

Put a ✓ next to the amounts that you are sure you **would** pay

Put a **X** next to the amounts that you are sure you **would not** pay

Put a circle around the **maximum amount** you would pay

£0

£5

£10

£15

£20

£25

£30

£35

£40

£45

£50

£55

£60

£65

£70

£75

£80

£85

£90

£95

£100 +

(If £100 + state the exact amount in the space provided below)

£ ____________________

4) **What is the maximum amount of money you would be willing to pay for <u>the one-off Flexi-scope test at the age of 65?</u>**

Put a ✓ next to the amounts that you are sure you **would** pay

Put a **X** next to the amounts that you are sure you **would not** pay

Put a circle around the **maximum amount** you would pay

£0

£5

£10

£15

£20

£25

£30

£35

£40

£45

£50

£55

£60

£65

£70

£75

£80

£85

£90

£95

£100 +

(If £100 + state the exact amount in the space provided below)

£ ____________________

5) **Could you please explain the reasons for your answers to questions 3 and 4?**
(please write your explanation in the space below)

GP ID CODE: gp1.1

6) We would now like to ask you some questions about yourself. To repeat, these details are entirely confidential.

Gender
(please tick the appropriate box)

- ☐ Female
- ☐ Male

Could you please tell us your age?
(please write your answer in the box) ☐ **years**

Ethnic origin:
Please tick the box against the term which describes you:

- ☐ White
- ☐ Black - African
- ☐ Black - Caribbean
- ☐ Black - Other
- ☐ Indian
- ☐ Pakistani
- ☐ Bangladeshi
- ☐ Chinese
- ☐ Asian - Other
- ☐ Other (please specify) ....

What is your marital status?
(please tick the appropriate box)

- ☐ single
- ☐ married/living with partner
- ☐ widowed

How many children do you have?
(please write your answer in the box) ☐

Are you currently:
(please tick the appropriate box)

- ☐ Employed
- ☐ Unemployed
- ☐ House-wife/house-husband
- ☐ Retired
- ☐ Student
- ☐ Other

If other please state: ....

Current or most recent occupation: ....

Are you the main income earner in the household?
(please tick the appropriate box)

- ☐ Yes
- ☐ No

If **NO**, what is the occupation of the main income earner? ....

Could you please estimate the annual income of your household before deducting tax and national insurance? (If you receive any benefits or pensions include them as income).
(please tick the appropriate box)

- ☐ Less than £10,000
- ☐ £10,000 - £20,000
- ☐ £20,001 - £30,000
- ☐ More than £30,000

What was your age when you left full-time education?
(please write your answer in the box)

☐

Have you or your close family ever suffered from any of the following health problems?
(please tick the appropriate box)

Yourself	Your family	
☐	☐	Stomach problems e.g. ulcer
☐	☐	Piles or haemorrhoids
☐	☐	Heart trouble
☐	☐	Cancer
☐	☐	Stroke
☐	☐	Severe depression

Would you say your health is:
(please tick the appropriate box)

- ☐ Excellent?
- ☐ Good?
- ☐ Fair?
- ☐ Poor?

Regarding smoking, would you describe yourself as :
(please tick the appropriate box)

- ☐ A smoker
- ☐ An ex-smoker
- ☐ Never smoked

How many times have you visited your GP in the last year?
(please tick the appropriate box)

- ❑ None
- ❑ Once
- ❑ Twice
- ❑ Three
- ❑ More than three times

How many times have you visited your dentist in the past two years?
(please tick the appropriate box)

- ❑ None
- ❑ Once
- ❑ Twice
- ❑ Three times
- ❑ Four times
- ❑ More than four times

What other diseases have you been screened for in the last 5 years?
(please tick the appropriate box)

- ❑ None
- ❑ Cervical cancer
- ❑ Breast cancer
- ❑ Other

  If other, please state: ....

In general, how important do you feel the following health measures are?
(please tick the appropriate box)

	Not at all	Somewhat	Moderately	Very	Extremely
to exercise regularly	❑	❑	❑	❑	❑
to eat plenty of fruit	❑	❑	❑	❑	❑
for a women to have a cervical smear at least once every 5 years	❑	❑	❑	❑	❑
for a women to have a breast screen every 3 years	❑	❑	❑	❑	❑

How worried are you about getting bowel cancer?
(please tick the appropriate box)

Not worried	A bit worried	Quite worried	Very worried
❑	❑	❑	❑

Compared with other men and women of your age, do you think your chances of getting bowel cancer are:
(please tick the appropriate box)

Much lower	A little lower	About the same	A little higher	Much higher
❑	❑	❑	❑	❑

Although screening for colorectal cancer is not routinely available in the UK, some people may already have already received an FOB or Flexi-scope test. If you have please indicate below.

❑ FOB
❑ Flexi-scope

WE WOULD LIKE TO TAKE THIS OPPORTUNITY TO REMIND YOU THAT YOUR ANSWERS ARE CONFIDENTIAL AND THAT THEY WILL NOT BE USED IN SETTING CHARGES FOR SUCH SERVICES

PLEASE PLACE THE QUESTIONNAIRE IN THE ENVELOPE PROVIDED

MANY THANKS FOR YOUR VALUABLE CONTRIBUTION TO OUR STUDY

Appendix Six

YOUR OPINION ON HEALTH SCREENING

Screening For Colorectal Cancer

Cancer of the colon or rectum - commonly referred to together as colorectal or bowel-cancer is the second most common cancer in the UK. It represents one of the most significant causes of cancer deaths in industrialised countries.

However, there are now grounds for believing that a national screening programme aimed at detecting colorectal cancer at a very early stage could save lives. There are several different screening techniques available for detecting colorectal cancer, these are currently being assessed in the UK.

We are trying to get an idea about how people would value a national screening programme for colorectal cancer and would appreciate your opinion. This project is being sponsored by the Medical Research Council and your answers will be used for research purposes only.

In order to help us, we would be grateful if you could take the time to complete the enclosed questionnaire. You do not have to do so, and your care will not be affected in any way if you decide not to take part in this survey.

There are no right or wrong answers to the questions asked. We are interested in your view. Please take time to read the descriptions of the disease and the screening investigations before attempting to answer the questions.

Some questions will ask you to tick a box like this ☑. Other questions will ask you to write your answer in the space provided. Please use black or blue ink.

This questionnaire is anonymous.
Your answers will be treated as confidential.
If this questionnaire causes concern, please contact Jane Wolstenholme or Emma Frew at the address below, or your GP. Complaints about the way the study is being conducted should be addressed to your Local Authority Complaints Department, (address{please see attached sheet}).**

Contact name & address:
Jane Wolstenholme or Emma Frew
Trent Institute for Health Services Research
Room B39, Medical School,
QMC,
Nottingham. NG7 2UH

tel: 0115 9709765

A Questionnaire to Assess the Value to You of Screening for Colorectal Cancer.

Colorectal Cancer:
Cancer of the colon or rectum - commonly referred to together as colorectal or bowel cancer is the second commonest cancer in the UK.

Why screen for colorectal cancer?
Screening means doing simple tests to pick up hidden problems. If people wait for symptoms before seeking care, there will be a high proportion of more advanced cancers detected, which are more difficult to treat. Patients whose cancers are detected at earlier stages by screening do much better than patients with more advanced cancers.

What does screening involve?
There are two main types of screening test for bowel cancer, neither are available in the UK at present. Trials are being carried out to assess whether it would be worth introducing a national screening programme for colorectal cancer using one of the screening tests outlined below.

A) Faecal Occult Blood (FOB) test .
The faecal occult blood test would involve screening once every two years from the age of 50 to 74 years old. The test is carried out by yourself at home and involves a small kit which would be sent to you in the post. It requires small stool (faeces) samples to be taken over a time period of three days, placed on special paper within the test kit and then returned to your local GP. Investigation of the test kit would enable the confirmation of the possible presence or absence of bowel cancer. If the test confirms the possible presence of cancer, you would be required to come into your local hospital for further investigations.

B) Flexi-scope test
The Flexi-scope screening test would involve a one-off screen for all people aged 60-65 years old. The flexi-scope is a thin flexible tube with a tiny camera on the end. A doctor inserts the Flexi-scope into the back passage and looks for bowel polyps. Most bowel polyps are harmless growths, but some of them may become cancerous. The majority of polyps can be removed quickly and painlessly with the Flexi-scope. Some of the polyps may need to be removed using a colonoscope under general anaesthetic. Removing these polyps helps to prevent bowel cancer. The Flexi-scope would be done at a local hospital clinic, no anaesthetic is required. An enema would be provided for you to use at home before the test to clear your bowel. The test would only take 5 minutes although it would be necessary to devote an hour to the visit because of waiting time. If an abnormality is found this can be often removed using the Flexi-scope during the same visit.

1) Would you have one of these screening tests for bowel cancer if it was available?
(please tick the appropriate box)

Yes	No	Don't know
❑	❑	❑

IF YOU TICKED 'YES' OR 'DON'T KNOW', PLEASE GO TO **QUESTION 2**

IF YOU TICKED 'NO', PLEASE GO TO **QUESTION 7**

2) Which type of method of screening would you <u>prefer</u> if it was available to you?
(please tick <u>one</u> appropriate box)

Faecal Occult Blood Test	Flexi-scope	No preference
❑	❑	❑

One way of measuring the value of screening for colorectal cancer is to ask you what you would be prepared to give up to receive this service i.e. how much money you would be willing to pay for it. Of course, if colorectal cancer screening did come into existence in England, the screening test would be provided <u>free</u> by the NHS. **<u>We also believe that people should not have to pay for health care This is simply a method of measuring how strongly you feel about having a new screening programme and how much you would value such a service.</u>**

3) Do you think having the <u>complete series of FOB tests every 2 years from the age of 50 until the age of 74</u> is worth:

£1000

(please tick the appropriate box)

☐ Yes ☐ No

4) Could you please explain the reasons for your answer to question 3?

5) Do you think having the <u>one-off Flexi-scope test at the age of 65</u> is worth:

£1000

(please tick the appropriate box)

☐ Yes ☐ No

6) Could you please explain the reasons for your answer to question 5?

7) We would now like to ask you some questions about yourself. To repeat, these details are entirely confidential.

Gender
(please tick the appropriate box)

- ❑ Female
- ❑ Male

Could you please tell us your age? ☐ **years**
(please write your answer in the box)

Ethnic origin:
Please tick the box against the term which describes you:

❑ White	❑ Pakistani
❑ Black - African	❑ Bangladeshi
❑ Black - Caribbean	❑ Chinese
❑ Black - Other	❑ Asian - Other
❑ Indian	❑ Other (please specify) ....

What is your marital status?
(please tick the appropriate box)

- ❑ single
- ❑ married/living with partner
- ❑ widowed

How many children do you have? ☐
(please write your answer in the box)

Are you currently:
(please tick the appropriate box)

- ❑ Employed
- ❑ Unemployed
- ❑ House-wife/house-husband
- ❑ Retired
- ❑ Student
- ❑ Other

  If other please state: ....

Current or most recent occupation: ....

Are you the main income earner in the household?
(please tick the appropriate box)

- ❑ Yes
- ❑ No

If **NO**, what is the occupation of the main income earner?..

Could you please estimate the annual income of your household before deducting tax and national insurance? (If you receive any benefits or pensions include them as income).

(please tick the appropriate box)

❑ Less than £10,000

❑ £10,000 - £20,000

❑ £20,001 - £30,000

❑ More than £30,000

What was your age when you left full-time education?

(please write your answer in the box) **years**

Have you or your close family ever suffered from any of the following health problems?

(please tick the appropriate box)

Yourself	Your family	
❑	❑	Stomach problems e.g. ulcer
❑	❑	Piles or haemorrhoids
❑	❑	Heart trouble
❑	❑	Cancer
❑	❑	Stroke
❑	❑	Severe depression

Would you say your health is:

(please tick the appropriate box)

❑ Excellent?

❑ Good?

❑ Fair?

❑ Poor?

Regarding smoking, would you describe yourself as :

(please tick the appropriate box)

❑ A smoker

❑ An ex-smoker

❑ Never smoked

How many times have you visited your GP in the <u>last</u> year?
(please tick the appropriate box)

- ❑ None
- ❑ Once
- ❑ Twice
- ❑ Three
- ❑ More than three times

How many times have you visited your dentist in the past <u>two</u> years?
(please tick the appropriate box)

- ❑ None
- ❑ Once
- ❑ Twice
- ❑ Three times
- ❑ Four times
- ❑ More than four times

What other diseases have you been screened for in the last 5 years?
(please tick the appropriate box)

- ❑ None
- ❑ Cervical cancer
- ❑ Breast cancer
- ❑ Other
  If other, please state: ....

In general, how important do you feel the following health measures are?
(please tick the appropriate box)

	Not at all	Somewhat	Moderately	Very	Extremely
to exercise regularly	❑	❑	❑	❑	❑
to eat plenty of fruit	❑	❑	❑	❑	❑
for a woman to have a cervical smear at least once every 5 years	❑	❑	❑	❑	❑
for a woman to have a breast screen every 3 years	❑	❑	❑	❑	❑

How worried are you about getting bowel cancer?
(please tick the appropriate box)

Not worried	A bit worried	Quite worried	Very worried
❑	❑	❑	❑

Compared with other men and women of your age, do you think your chances of getting bowel cancer are:
(please tick the appropriate box)

Much lower	A little lower	About the same	A little higher	Much higher
❑	❑	❑	❑	❑

Although screening for colorectal cancer is not routinely available in the UK, some people may already have already received an FOB or Flexi-scope test. If you have please indicate below.

❑ FOB
❑ Flexi-scope

8) Have you recently completed a similar colorectal cancer screening questionnaire?
(please tick the appropriate box)

❑ Yes ❑ No

WE WOULD LIKE TO TAKE THIS OPPORTUNITY TO REMIND YOU THAT YOUR ANSWERS ARE CONFIDENTIAL AND THAT THEY WILL NOT BE USED IN SETTING CHARGES FOR SUCH SERVICES

PLEASE PLACE THE QUESTIONNAIRE IN THE ENVELOPE PROVIDED

MANY THANKS FOR YOUR VALUABLE CONTRIBUTION TO OUR STUDY

PAGE/PAGES EXCLUDED UNDER INSTRUCTION FROM UNIVERSITY