To Doriana and her passion for life that endures in all of us

Several people helped me through this undertaking and deserve my thanks. They include my friends in Milan and London, my fellows at Bocconi University, my uncle Giovanni and my parents who emotionally supported me the whole time.

Four people have really made the difference and therefore deserve full recognition for their role in preparing this dissertation. Jacopo, born three years ago, gave me the optimism and joy required to find the pleasure to study. I wish I had spent more time with him, but I am sure that his passion for baby books owes something to LSE. I met Giuliana, my love and wife, in London and thus I am thankful to my PhD experience also for this. She has always encouraged me, despite the trade-off between studying and our life as a couple. Without her I would have never found the emotional balance to commit myself to this enterprise.

Writing a PhD is mainly an individual experience. However, the supervisor makes this experience a bit less lonely and has a tremendous impact on motivation. Julian Le Grand was the best supervisor I could find, as he has always been supportive and motivating, in addition to providing me with clear and rich intellectual direction. Elio Borgonovi, who has supported me in many respects for 20 years now, is the person who mostly influenced my professional and intellectual life. He strongly encouraged me to enrol at LSE and has always helped me to proceed with my studies. Without his tenacity, which enhanced mine, and his vision I would not have written this thesis.

A Cost-Benefit Analysis of In-Vitro-Fertilisation in Italy

Giovanni Fattore

The London School of Economics and Political Science

Thesis submitted to the University of London for the degree of Doctor of Philosophy

2008

Abstract

This thesis presents a full cost-benefit analysis of In-Vitro-Fertilisation (IVF) from a societal perspective. It is based on a contingent valuation survey administered through internet to a sample of the Italian population. A referendum format and a payment scale were used to elicit willingness-to-pay (WTP) for a publicly funded program providing IVF to infertile couples. WTP was also elicited for a hypothetical situation in which the respondent was asked to imagine being infertile and willing to have a baby.

Overall, results show the feasibility of using this new method of administration of contingent valuation questionnaires. Responses reveal consistent patterns and the number of inconsistent answers is limited. WTP for private use (in case of infertility) and for a public program are positively associated with income, education, being within the fertility age range and being informed about infertility and IVF.

The take-it-or-leave-it format and a variant of the payment scale method result in different mean WTP estimates, but simulated and actual referendum WTP are very similar. There is evidence of an anchoring effect since the values presented in the take-it-or-leave-it question had an impact on the answers to the modified payment card questions that followed.

Mean WTP estimated from the different questions are consistently above the mean cost of providing IVF, as estimated on the basis of a full costing methodology. The IVF program shows net welfare benefits under several assumptions. The study shows that the societal benefits of an IVF program mainly derives from the high WTP of a minority of citizens who tend to be the most affluent, educated and familiar with infertility and IVF.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AR	Assisted Reproduction
ART	Assisted Reproduction Techniques
ACE	Cost-effectiveness Analysis
ACU	Cost-Utility Analysis
CBA	Cost-Benefit Analysis
CV	Contingent Valuation
DC	Dichotomous Choice
DCE	Discrete Choice Experiment
IVF	In-Vitro-Fertilisation
MPC	Modified Payment Card
NICE	National Institute for Clinical Excellence
NOAA	National Oceanic and Atmospheric Adminisration
OLS	Ordinary Least Squares
OE	Open-ended
OR	Odds Ratio
PC	Payment Card
SH	Structured Haggling
TIOLI	Take-it-or-leave-it
WTA	Willingness-To-Accept
WTP	Willingness-To-Pay

Chapter 1

Introduction

1.1. The purpose of the study

The aim of this thesis is to present a full cost-benefit analysis of In-Vitro-Fertilisation (herein referred to as IVF). The research was designed and implemented to offer guidance on policy making for public funding of this form of intervention. Basically, we wanted to shed some light on the economic rationality for using government resources to provide IVF, a relatively sophisticated and expensive procedure that increases the chances of infertile couples of having a baby. Therefore our study deals with the allocation of resources in public health care systems and aims at illustrating how cost-benefit analysis may concretely help in making decisions about rationing.

The concepts of rationing and priority setting applied to healthcare are now receiving greater attention in many industrialised countries. In the Netherlands, Italy, New Zealand, Norway, Sweden, United Kingdom and the United States of America various initiatives have openly recognised that diverse forms of rationing take place in healthcare systems and that, at least potentially, there are merits in taking some steps towards more explicit, systematic and democratic rationing.

Although it is increasingly recognised that implicit rationing is becoming less sustainable, so far only a few concrete initiatives have made rationing more explicit. Very often the policy debate tends to remain on the ground of principles and general criteria. However, economic evidence is required for reimbursement or coverage of pharmaceuticals in many countries (Taylor et al., 2004). The National Institute of Health and Clinical Excellence (NICE) in England and Wales and the Scottish Medicines Consortium (SMC) in Scotland regularly produce guidance based on economic evidence (Cairns, 2006). In Italy, the Netherlands, Canada, Australia and Sweden economic evidence is often required by regulatory bodies, but the process according to which economic data are analysed and assessed appear less well established than in the United Kingdom. Over the last ten years there have been a significant moves towards making governments' decisions about funding medicines and, to a lesser extent, other technologies and interventions based on evidence of effectiveness and cost-effectiveness.

Despite the advances in explicit and systematic approaches to rationing and a wider use of economic evaluation studies to guide policy making, we are aware of a few explicit exclusions from public coverage in Europe. Despite clear evidence that resources made available by governments do not suffice to meet demand, rationing decisions are often left at organisational level and at the point of service delivery (Klein et al., 1996).

Nevertheless, there are some services that represent exceptions, as they tend to be a preferred target for explicit exclusion from public coverage - IVF is one of these exceptions. In 1994, almost a quarter of the Health Authorities in Britain refused to purchase IVF services (Evans, 1995). President Clinton's ill-fated Health Security Act specifically excluded IVF services from the minimum package of services to be covered by insurance plans. In the Netherlands, the Dunning Committee (a committee set up to define the legal responsibilities of insurance schemes), argued that IVF should not be considered necessary, since childlessness does not interfere with the normal functioning of Dutch society.

Despite IVF having been targeted by explicit rationing in several situations, little is known about what IVF is worth to patients and to society. In particular, very few studies have been carried out to measure the value generated by the money spent on IVF treatment. This study aims to investigate the value generated by resources used on this intervention, which increases the probability of couples with fertility problems to have a child.

The concepts of rationing and priority setting applied to healthcare are now receiving greater attention in many industrialised countries. In the Netherlands, Italy, New Zealand, Norway, Sweden, United Kingdom and the United States of America various initiatives have openly recognised that some forms of rationing take place in the healthcare systems and that, at least potentially, there are merits in taking some steps towards more explicit, systematic and democratic rationing.

Although it is increasingly recognised that implicit rationing is becoming less sustainable, so far only a few concrete initiatives have made rationing more explicit. Very often the policy debate tends to remain on the ground of principles and general criteria. However, economic evidence is required for reimbursement or coverage of pharmaceuticals in many countries (Taylor et al., 2004). The National Institute of Health and Clinical Excellence (NICE) in England and Wales and the Scottish Medicines Consortium (SMC) in Scotland regularly produce guidance based on economic evidence (Cairns, 2006). In Italy, the Netherlands, Canada, Australia and Sweden economic evidence is often required by regulatory bodies, but the process according to which economic data are analysed and assessed appear less well established than in the United Kingdom. Over the last ten years there have been a significant moves towards making governments' decisions about funding medicines and, to a lesser extent, other technologies and interventions based on evidence of effectiveness and cost-effectiveness.

Despite the advances in explicit and systematic approaches to rationing and a wider use of economic evaluation studies to guide policy making, we are aware of a few explicit exclusions from public coverage in Europe. Despite clear evidence that resources made available by governments do not suffice to meet demand, rationing decisions are often left at organisational level and at the point of service delivery (Klein et al., 1996).

Nevertheless, there are some services that represent exceptions, as they tend to be a preferred target for explicit exclusion from public coverage - IVF is one of these exceptions. In 1994, almost a quarter of the Health Authorities in Britain refused to purchase IVF services (Evans, 1995). President Clinton's ill-fated Health Security Act specifically excluded IVF services from the minimum package of services to be covered by insurance plans. In the Netherlands, the Dunning Committee (a committee set up to define the legal responsibilities of insurance schemes), argued that IVF should not be considered necessary, since childlessness does not interfere with the normal functioning of Dutch society.

Despite IVF having been targeted by explicit rationing in several situations, little is known about what IVF is worth to patients and to society. In particular, very few studies have been carried out to measure the value generated by the money spent on IVF treatment. This study aims to investigate the value generated by resources used on this intervention, which increases the probability of couples with fertility problems to have a child.

1.2. Why Cost-benefit Analysis?

As argued by New and Le Grand (1996) moving towards more explicit, systematic and democratic forms of rationing is fraught with danger. In the areas where there is explicit rationing, decisions are made on the basis of facts rather than hypotheses. There is a clear need to identify rational approaches. One way to offer guidance to decision makers on a specific rationing decision is to perform an economic evaluation. In this type of evaluation, healthcare decisions are assessed by comparing costs of alternatives with their consequences or outcomes (Drummond et al., 1997). Traditionally, economic evaluations in the healthcare field have been carried out as either cost effectiveness-analysis or cost-utility analysis (ACU). In cost-effectiveness analysis (ACE) the cost per unit of health effects (i.e. life-years) gained by the adoption of the programme is estimated. In cost—utility analysis health effects are measured in terms of utility, in order to capture and combine in a single measure the different health benefits according to individuals' judgement. Both cost-effectiveness and cost-utility analysis are inappropriate in the context of IVF. While it is theoretically possible to estimate the cost per new life or even the cost per Quality Adjusted Life Year (QALY) attributable to this technology, it is likely that these approaches will miss the real benefit of IVF: is the increase in welfare of giving an infertile couple a chance to have a baby.

Cost-effectiveness analysis may provide important economic insights about IVF and other Assisted Reproduction Technologies (ARTs). The incremental cost-effectiveness ratios, where effectiveness is measured in terms of live births or pregnancies, may be used to make a preliminary assessment about the "value" of the intervention and can be used to compare different ARTs or different categories of patients undergoing ARTs. Indeed, later in this thesis we present a cost-effectiveness analysis of IVF and use it to make sense of the data obtained in the contingent valuation (CV) survey. However, the cost per live birth of using IVF or other ARTs cannot be compared to the cost per life saved or per year of life gained. Simply, new babies and saved lives are not commensurable as they refer to two completely different categories of benefits. Despite that both concern with life, there are no reason to assume that the contribution to the wellbeing of a new life is similar to that of a saved life. Therefore, cost-effectiveness data provide weak evidence, if any, to make decision about public coverage of IVF.

Cost-utility of healthcare intervention is becoming increasingly popular. In this type of study, health consequences of interventions are evaluated according to indexes (Quality Adjusted Life Years to mention the most widespread) that should represent a summary of patients' preferences for different health states. In theory, the QALY construct could be used to measure the intensity of preference of a person (or a couple) for having a child in the case of infertility. For example, standard gamble exercises could be designed to ask respondents to trade between their years of life and the probability of having a child using IVF treatment. However, asking a person to trade her personal years of life with the probability of a newborn from IVF is logically far away from the QALY concept and the way it can be elicited.

An alternative approach to assess IVF from an economic perspective is to carry out a cost-benefit analysis (CBA). The main difference between CBA and the other techniques mentioned above is that the former seeks to place monetary values on both the inputs (costs) and outcomes (benefits) of the programme being evaluated, while the latter provides monetary measures of only the costs.

CBA is more appropriate than ACE and ACU because it allows one to capture the full value that IVF treatment holds for people. In the other two techniques, the value of programmes is strictly related to the contribution to the health improvements that the beneficiaries enjoy. Even in ACU, that makes use of the utility concept, the basic idea is to measure health-related quality of life rather than quality of life in general. Infertility is a medical condition; IVF is a medical intervention, being childless can be a very distressful condition - childlessness can sometimes lead to psychological disorders. Nevertheless, the gain that people derive from IVF is not mainly related to the health of the parents. It derives from fulfilling the desire of becoming a parent and its impact is on the overall wellbeing of the person/couple.

Cost-benefit analysis is more appropriate that ACE and ACU because it is more flexible and can thus go beyond the narrow health implications of procreating babies through IVF. It can measure the value of obtaining a child when one is infertile as well as also capture the benefits of undergoing the entire process, which can be associated with avoiding regret feelings (Ryan, 1999). In addition, CBA appears better grounded in welfare economics theory and provides clearer decision rules than ACE and ACU.

There are three steps in conducting a CBA for this specific medical intervention: a) to perform a cost-analysis to estimate the cost per cycle of treatment; b) to review effectiveness evidence to calculate the cost per unit of success, that is the cost per delivered baby; c) to compare the cost per delivered baby with the monetary value placed by society to have a baby from IVF. This study covers all the three steps and tries to provide some innovative advancements in terms of methodology on both costing and benefit measuring systems.

In order to measure the benefits of the treatment, one option is to refer to the human capital concept. According to this concept, the benefits of a health care intervention can be measured in terms of the future flow of income that is freed by the intervention (Robinson, 1993). The human capital approach has several limitations (Robinson, 1993; Johannesson 1996a). From a pragmatic point of view it has the disturbing consequence of assigning very small values to poor people and to individuals who are not in the labour force. From a theoretical point of view, the approach is not consistent with welfare economics because it is not rooted in the proper concept of willingness to pay (WTP), as measured by Kaldor and Hicks. Finally, at an intuitive level, the fallacy of the human capital approach derives from the fact that better health is measured in terms of enhanced productivity only, thus neglecting that individuals value good health per se. As a consequence, the human capital approach is no longer popular among economists and standard cost-benefit analysis is now generally carried out using the WTP approach. In this approach, monetary values are placed by observing or eliciting how much money individuals are prepared to pay for a particular good.

WTP can be measured using two main approaches: revealed preferences or expressed preferences. The revealed preference approach is based on the observation of actual behaviour, while the expressed preference approach is based on the direct elicitation of WTP from individuals through the use of carefully designed and administered sample surveys. This approach is usually named the contingent valuation (CV) method.

In this thesis we use the CV method to perform a full cost-benefit analysis of a government programme providing IVF to infertile couples. The CBA aims to provide solid evidence to policy makers. In particular, it aims to use economic analysis to inform decisions relating to government coverage of a major medical intervention that has been specifically targeted by explicit rationing decisions. More specifically, two main research questions are addressed by this empirical study:

1. Can a contingent valuation survey conducted via the internet be a feasible method to measure benefits of a health care programme in general, and of a programme providing IVF in particular?

2. Taking into account the experimental state of the methodology, is the expressed WTP of Italians for a public programme providing IVF to infertile couples greater than costs of implementing the programme? In other words, would the benefits of such programme exceed its costs?

These research questions are strongly connected with four broader issues that are discussed in the subsequent chapters. The first issue pertains to IVF. Through the survey we have collected evidence to broaden our knowledge concerning the attitude towards this particular service. IVF, despite absorbing relatively small amounts of resources, well represents the new medical technologies that raise exceptional ethical problems and in which treatments aim at improving human functioning and general wellbeing rather than curing diseases.

The second aspect concerns the relationship between economics and health policy. By investigating the feasibility of a specific economic tool we expect to better understand how economic "rationality" can establish a fruitful dialogue with the multitude of perspectives that influence decision making processes within the health care sector. The IVF case is analysed in the broader context of the problem of the scarcity of resources in the healthcare system. The specific focus of the empirical part of the thesis is on the feasibility of using economic analysis to offer guidance to rationing decisions.

Previous contingent valuation studies have mainly investigated the perspective of IVF users. Given that our research measured the willingness-to-pay (WTP) of a sample of the general population we were able to investigate the role of caring externalities. Therefore, the third aspect discussed in the thesis concerns the distinction between egoism and altruism in the context of reproductive medicine.

Finally, this thesis investigates a thorny issue in contingent valuation studies: the format to elicit WTP. Since our sample was asked three WTP questions we could compare the referendum to the Payment Card formats and investigate how they differ.

1.3. Structure of the thesis

A full understanding of a cost-benefit analysis of IVF requires some understanding about infertility, its treatment and a variety of issues that makes this medical intervention rather special. Chapter 2 deals with these issues thus presenting the background of the rest of the thesis. It also mentions some policy issues raised by the use of this intervention and its government funding.

In a full cost-benefit analysis both estimates of costs and benefits are important. Therefore we paid attention to both of these issues in the thesis. Chapter 3 reviews the literature on contingent valuation and specifically investigates previous contingent valuation studies on ARTs. The review of the vast literature review on CV in healthcare has provided us with important elements to design the survey and has been used to make sense of our results.

To provide a solid background to the costing part of the study we used a rather different approach. We went through the literature on cost accounting, generally considered a sub-discipline of management, and we tried to learn from this literature how to conduct a cost study for the purpose of a cost-benefit analysis aimed at providing guidance about public coverage of an intervention (chapter 4). Briefly, we reached the conclusion that cost-benefit analysis should be based on solid cost studies, rather than on the use of tariffs and prices, and that the full costing methodology has to be used.

Chapter 5 presents the methods used in the cost-benefit analysis and details the source of the data. The measurement of benefits is based on a large survey conducted via internet that was specifically designed for the purpose of this research. The survey collected information about the knowledge, attitude and willingness to pay for IVF of a sample of the Italian population. The sample is not probabilistic but it was built to make a representative picture of adult Italians with respect to age, gender, education and residence. To our knowledge, this is the first attempt to investigate WTP for a medical intervention through a CV survey administered via internet to a large national sample. The last part of chapter 4 presents the methods used to estimate the cost of IVF. It explains how we calculated the cost of an IVF cycle according to a full costing procedure and how we used it in the context of the CBA. The cost analysis presented in the thesis is based on data collected in two Assisted Reproduction Centres located in northern Italy, one run by the Italian National Health Service and the other private.

Presentation of results is split into two chapters. Chapter 6 reports the results of the survey to provide a picture of the information about infertility and IVF, attitude towards this technology and willingness-to-pay. The purpose of this chapter is to provide a picture of what the people in the sample think about these issues and to assess possible correlations. In this chapter, we present results of several regression analyses that try to explain how demographic and socio-economic variables influence knowledge, attitude and willingness to pay for IVF. In addition to improve our understanding of what people think about IVF, this chapter provides some important insights about the validity of the survey and the contingent valuation method.

Chapter 7 focuses on the economic results of the study *strictu sensu*. It presents the estimates of mean WTP for personal use of IVF in case of infertility on the basis of the data of the Payment Card questions. It also presents WTP for a national programme providing IVF to infertile couple with public funding on the basis of the answers to a referendum format question and to a second Payment Card battery of questions. Results are then compared, discussed and used to investigate their validity. The chapter then reports results of the cost analysis and briefly reports results of a cost-effectiveness analysis. Finally, estimates of benefits and costs are assembled to perform the cost-benefit analysis of a programme providing IVF to infertile couples in Italy. On the basis of reasonable assumptions, our study shows that such a programme would generate net benefit and thus should be endorsed by the government.

We conclude our thesis with a discussion of several methodological issues encountered and raised by the study (chapter 8). Overall, the study is encouraging as it shows that collecting information for cost-benefit analysis in population surveys is feasible and valid, at least to a certain extent. We do think that the spectacular advancements of Information and Communication Technologies create more opportunities to use public opinion to make collective choices, including public opinion data that can be used in the framework of rigorous economic analysis.

Overall, this study provides evidence in favour of the public funding of IVF. This funding is consistent with recent trends in many affluent countries where IVF to infertile couples has become part of the benefits of statutory coverage. Nevertheless, it is important to underline that the results of our study are mainly driven by a minority of respondents that have high willingness-to-pay for the programme. This study confirms that measurement of benefits in cost-benefit analysis may favour the point of view of those who are more affluent and thus have higher willingness to pay.

Chapter 2

Infertility and Assisted Reproduction Techniques

2.1. Introduction

This chapter provides a background to the rest of the thesis on infertility, the treatments to overcome it and some relevant ethical issues that influence policy making in this field of medicine. The chapter does not intend to cover in detail all the technical, ethical and legal aspects related to Assisted Reproduction Techniques (ARTs). Nevertheless it is written with the intention to give the reader a sufficient background to understand the rest of the thesis and to contextualise the use of economic analysis to offer guidance on public coverage of In Vitro Fertilisation (IVF) services.

The first two sections define infertility and its main treatments and provide a brief history of IVF, which was pioneered in the UK in the late 1970s. The following two sections concisely report the available evidence on the effectiveness and the possible side effects of IVF: it is now clear that this treatment does help infertile people to conceive but also presents a serious clinical side effect that is multiple births.

Section six briefly reports the main ethical problems at stake with ARTs. The aim of this section is to provide the reader with a summary of the issues that can influence public opinion and, in particular, attitudes towards public funding. The ethical and social issues arisen by IVF and other ARTs motivated European governments to enact specific legislation to regulate the matter. A summary of the regulation and public funding of IVF in the major EU member states are reported in section 7 and 8.

2.2. Infertility and its treatments

Infertility is generally defined as the inability to conceive after 12 months of unprotected sexual intercourse (Udoff and Adashi, 1999). However, the National Institute for Clinical Excellence (NICE) in the UK suggests a more stringent definition:

that is, failing to get pregnant after two years of regular unprotected sexual intercourse. Infertility can be classified as either primary or secondary. Primary infertility is diagnosed if the couple has never had a conception, while secondary infertility is the term applied to couples that have a history of at least one documented conception but who currently have not been able to conceive for at least twelve months (if the less stringent definition of infertility is applied).

It should be noted that the definitions of infertility are based on evidence of pregnancy over a period of time, rather than on a physiological status of the couple. The probability of a normal couple conceiving in a given menstrual cycle is called fecundability and is estimated to be 25%. Taking into account that a fraction of conceptions results in spontaneous abortion, it is estimated that over 90% of normal couples will have conceived at the end of one year of unprotected sexual intercourse. Therefore, for those who fail to conceive the diagnosis of infertility (as defined above) does not necessarily mean that conception cannot occur, but rather that the couple have been unable to conceive naturally over a twelve month period. These couples belong to a medically defined group that is less likely to conceive. Some of them may be able to conceive naturally after a few more attempts while others may be affected by various conditions that make natural conception either highly unlikely or impossible.

Infertility is thought to affect 10-15% of all couples of reproductive age. Infertility increases with the age of both the woman and the man. The aetiology of infertility can be divided into three major categories: (i) male factors, (ii) female factors and (iii) undetermined aetiology. Men and women equally contribute to couple infertility. In approximately 40% of couples the aetiology is primarily male factor. Female factor accounts for another 40% while the remaining 20% of couples' infertility is attributed to a combination of male and female factors.

The investigation of an infertile couple is a rather long process beginning with the analysis of a complete medical history and including a physical exam and laboratory testing. The aim of the evaluation is to make a specific diagnosis, which is helpful to decide the appropriate course of treatment. For male infertility approximately 50% of cases have no specific aetiology and for 75% there is no treatment that can directly improve the abnormality. The advent of Assisted Reproductive Technologies (ARTs) greatly benefited couples affected by male infertility as in almost all cases it can overcome the disfunction.

In approximately 40% of infertile women, the main aetiological factor is the failure to ovulate and in another 40% is related to tubal damage or other pelvic pathologies. Unusual causes and unexplained infertility account for the remaining 20%.

Treatment strategies for infertility depend on the specific aetiology of the infertility. The main strategies are surgery for anatomical defects and other specific conditions, ovulation induction with various drugs including compounds made through recombinant technology and ARTs, including sperm or egg donation. Indications for ARTs include tubal factor infertility, endometriosis, male factor infertility, immunologic infertility, and unexplained infertility. IVF is not the only or the most important intervention to overcome infertility. It is important to bear in mind that more simple techniques like artificial insemination are frequently used. Furthermore, it should be noted that pharmaceutical therapies increase the probability of pregnancy in many categories of patients. Scientists and clinicians suggest following a gradual approach to treating infertility and recommend the use of complicated (and often invasive) techniques only if the more simple ones fail.

IVF consists of a sequence of steps that, by overcoming possible male and female dysfunctions, make conception more likely than through the natural way. In the natural cycle an oocyte (egg) leaves an ovary and is transported through a fallopian tube. There it can be fertilized by a spermatozoon and become an embryo. The embryo may settle in the uterus and grow into a foetus. In an IVF cycle this sequence is altered by medical intervention. First, medication is given in order to obtain more oocytes. Then, if stimulation is successful, the oocytes are aspired from the ovaries. Here the in vitro phase begins. Basically, spermatozoa (previously collected) and oocytes are brought together in order to obtain embryos. These embryos are then transferred into the woman's uterus three or five days after the aspiration. From this stage on, natural and assisted conceptions return to follow the same course.

ZIFT (Zygote intra-Fallopian transfer), GIFT (Gamete intra-Fallopian transfer) and ICSI (Introcytoplasmic injection) are variants of IVF. ZIFT is the same procedure as

IVF except that embryos are transferred into the Fallopian tubes rather than into the uterus. GIFT utilizes the same oocyte stimulation protocol as IVF; however, once the eggs are retrieved, they are mixed with sperm and immediately transferred into the tubes. GIFT and ZIFT techniques are not widely used and, according to the NICE Clinical Guideline issued in 2004, there is insufficient scientific evidence to recommend their use in preference to conventional IVF (National Collaborating Centres, 2004). ICSI is a micromanipulation technique. This sophisticated procedure introduces a single sperm directly into the cytoplasm of the egg and can produce fertilization even in cases with very poor-quality sperm. This technique overcomes almost all the causes of male infertility (Edwards, 1998), improves fertilisation rates and represents more than 50% of all IVF treatment in many countries (HFEA, 2006).

2.3. IVF and infertility treatments: a brief history

For centuries, human reproduction was a mystery and, in many societies, birth was considered as a miracle. Those who were unable to conceive could not receive significant help and were often stigmatised. Until the beginning of the last century, a variety of rites and primitive medicines, including ingredients such as pig's teeth, frogs and spiders were the only hope for infertile couples (Edwards, 1998). Indeed, infertility, up until recently, was believed to be a female problem, although it is now undisputed that both female and male infertility cause couple infertility (see above).

IVF has diverse origins (Edwards, 1998) and the history of the technique may be characterised by three related periods of development (Iglesias, 1990). The first phase dates from the late nineteen century until the late 1960s - it was in the hands of research scientists. Dr. Heape is the person generally attributed as being the first researcher to make a significant contribution to the development of IVF. Heape transferred genetically marked embryos from one rabbit to another and obtained offspring. In the thirties the first attempts were made to culture mammalian embryos in vitro. Approximately during the same period various hormones influencing and regulating the gonads were discovered and in the following two decades they were tested on mice to induce ovarian stimulation.

The second phase, from 1968 to 1978, is characterised by the collaboration between scientists and clinicians. In 1968 the first medico-scientific research team was set up under the leadership of Dr. Edwards (a physiologist who is considered the father of

IVF) in England. The team set up a very ambitious agenda: "to attempt (human) fertilisation in vitro and the culture of embryos in order to alleviate some forms of infertility, and to study the origin of inherited defects" (Edwards, 1983). In effect, this ten-year period led to the acquisition of all the fundamental steps required to implement IVF. They included the basic understanding and control of the maturation of a) the human egg, follicle growth and the process of ovulation, b) the fertilising capacitation of sperm, c) the aspiration of the pre-ovulatory egg, d) the replacement of the embryo in the womb, d) the implantation on the embryo in the womb and the continuation of pregnancy (Iglesias, 1990). Mastering these steps has made it possible to replace embryos into infertile mothers since 1972. The birth of the first baby successfully conceived in vitro, Louise Brown, was in July 1978.

The third phase (1978 to present), has seen significant technological developments and very rapid widespread use of the technique in industrialised countries. In the 1980s new drugs were introduced, the IVF technique was refined and variants of IVF were developed (gamete intra-fallopian transfer –GIFT- and zygote intra-fallopian transfer – ZIFT). In 1986 the diagnosis of genetic disease in human pre-implantation embryos was introduced. More recently, ICSI (intracytoplasmic sperm injection) has opened the era of direct manipulation of the human gamete. ICSI has significantly improved assisted reproduction by improving fertilisation rates in general and by extending the treatment population to include patients with male sub-fertility - previously not considered viable for classic IVF.

Since their first clinical success, IVF and other assisted reproductive techniques were rapidly adopted in many countries. At Bourn Hall, the private clinic established in England by the pioneers of IVF, from October 1980 to April 1983 139 babies were born as a result of IVF. In 2000 in the United Kingdom (UK), 72 IVF clinics were operating and about 6,500 babies were born after IVF or ICSI cycles (HFEA, 2000). Between April 2003 and March 2004 10,242 children were born through IVF (HFEA, 2006). They are approximately 1.5% of all the newborns in UK. These babies were procreated as a result of the commitment of 30,000 patients who underwent IVF for a total of about 38,000 cycles of IVF treatment.

A survey conducted by Shenker (1997) shows that in 1995/96 IVF was available in 516 centres in 39 European countries. According to a similar survey carried out on

Asian centres, approximately 267 ART centres (half of which in Japan) were operating in Asia (Shenker and Shushan, 1996). However, these centres were concentrated in only 16 countries, where the standard of living is relatively high. In the United States of America, where IVF was introduced in 1982, the market for IVF, excluding fertility drugs, is estimated at 1 trillions US $ (Spar, 2006). In 1986, there were roughly 100 fertility clinics, performing about 10,000 cycles of ARTs. In 2002, there were 428 clinics performing as many as 115,000 cycles.

2.4. Effectiveness of IVF

Despite the impressive development of IVF, the documentation of its effects has not been supported by large randomised clinical trials. Effectiveness is mainly documented by success rates reported by clinics and by national and international registries. Only recently, a few relatively small trials have reported significantly higher live birth rates with IVF/ICSI cycles when compared with no treatment in some categories of patients (National Collaborating Centres, 2004). Despite the lack of large clinical trials, however, it appears now undisputable that IVF is an effective procedure.

Since IVF is a chain of several treatment phases, success rates can be presented in several ways. The two main most meaningful measures are the pregnancy-per-cycle rate and the live birth-per-cycle rate. The pregnancy-per-cycle rate is the percentage of IVF cycles started that produced a pregnancy. This rate is higher than the live birth per cycle rate because some pregnancies end in miscarriage, therapeutic abortion, or stillbirth.

The live birth-per-cycle rate is the percentage of IVF cycles started that result in a live birth (a delivery of one or more babies). In terms of overall benefits, this is the most significant rate as it represents the average chance of having a live-born infant by using IVF. This rate is the main measure that will be used in this thesis.

The first significant statistics collected in the second half of the '80s reported a live birth-per-cycle rate around 10% (Haan, 1991; Medical Research International, 1989). Since then the success rate has significantly improved. In the early '90s patients undergoing IVF could expect pregnancy rates of 17% to 23% per cycle and corresponding delivery rates of 13 to 18% per cycle. Data from 49 countries for 2002 show that the delivery rate per initiated cycle is now 18.6% for conventional IVF and 20.4% for ICSI – the data refers to non donated material only (Adamson et al., 2006).

More recent data appear even more favourable. In the USA the 2003 Assisted Reproductive Technology Report shows basic statistics of almost all the clinics (399 of the 429 clinics) operating in the USA in that year (Centers for Disease Control and Prevention, 2005). Results reported in table 1 refer to all ART (IVF, GIFT and ZIFT) cycles that used fresh, non-donor eggs or embryos. Results mainly refer to IVF as GIFT and ZIFT were performed rarely (about 0.5% of treatments), while 56% of IVF cycles were performed using the micromanipulation technique ICSI (success rates in IVF cycles with and without ICSI appear similar). In 2003, according to the Report, 91,032 cycles were initiated with fresh non-donor eggs or embryos. They were responsible for 31,348 pregnancies and 15,367 live births. Consequently, pregnancy rates and live birth rates were 34.4% and 28.3%, respectively.

Table 2.1. Outcome of ART cycles using fresh, non-donor eggs or embryos in 2003 in the United States

	#		
Cycle initiated (1)	91,032	Pregnancy per cycle (4/1)	34,4%
Egg retrieval (2)	79,602	Live birth per cycle (5/1)	28,3%
Embryo transfers (3)	74,296	Live birth per retrieval (5/2)	32,4%
Pregnancies (4)	31,348	Live birth per transfer (5/3)	34,7%
Live births (5)	25,775	Live birth per pregnancy (4/5)	82,2%

Source: Centers for Disease Control and Prevention, 2005.

A similar registry is kept in the United Kingdom by the Human Fertilisation and Embriology Authority (HFEA, 2006). During the period 2003/04 the Authority collected data on 38,264 IVF cycles administered to 29,688 patients (the all cycles legally performed in the UK). In the 92/93 period there were reported 17,301 initiated cycles with a live birth per initiated cycle of 13.2%. In the late nineties the number of cycles performed in the UK almost doubled (28,689) and live birth rated reached 19.6. Data based on treatments carried out between April 2003 and March 2004 show that about 29,700 patients underwent IVF, that 38,264 cycles were initiated and that they facilitated 8,251 successful births (success rate = 21.6%).

There are various factors that are supposed to affect IVF pregnancies and live birth rates. The most significant one for the purpose of this dissertation is the age of the woman. It is now clearly recognised that the woman's age has a crucial effect on IVF success rate – this is clearly demonstrated by the American and the British registries. The evidence illustrates that pregnancy rates, live birth rates for ART Cycles using fresh non-donor eggs or embryos decrease with age. In the USA, while live birth rates for women who are 30 years old or less is above 40%, the average chance of live birth is 15.1% for women aged 40 (Centers for Disease Control, 2005).

The strong impact of age on IVF success rate is also reported in the UK (HFEA, 2006). Table 2 shows live birth rates by woman age. The live birth rates remain above 28% for women less than 35 years old. For older women this rate sharply declines, especially for those who are above forty. Out of 836 cycles performed in women aged 43 or above only 3.2% of the treatments resulted in a live birth. As we will see later in this thesis, these data have a very strong impact on the cost-effectiveness of IVF: older women, who are also often those who feel more compelled to procreate, present much less favourably in the cost-effectiveness ratios due to the limited effectiveness of the procedure on them.

Table 2.2. Live birth rates by age of woman in the United Kingdom (non-donor fresh eggs only)

Woman age	Treatment cycles	Live birth rate per treatment cycle
Under 35	13,489	28.2%
35-37	7,077	23.6%
38-39	3,984	18.3%
40-42	2,965	10.6%
43 and over	836	3.2%
All patients	28,351	23.3%

Source: HFEA, 2006

2.5. Clinical side effects of IVF: the problem of multiple births

Although IVF and other ARTs have brought hope to many couples suffering from infertility, these procedures have important shortcomings (Serour et al., 1999). The whole experience that couples have to undergo whilst on IVF programmes is emotionally taxing and stressful. From the clinical point of view, various side effects are associated to the procedure itself. Daily injections are required, tests have to be repeatedly carried out to monitor hormone levels, furthermore oocyte recovery and embryo transfer have some relevant side effects. The side effects that can result can be as a consequence of the medication; it increases the risk the Ovarian

Hyperstimulation Syndrome, as well as the risk of infection and bleeding during oocye collection (Serour et al., 1999). Moreover, women undergoing IVF experience higher rates of miscarriage, ectopic pregnancy and heterotopic pregnancy. However, it is unclear if the higher rates can be attributed to the interventions or simply to the baseline condition of the woman (for example some of these side effects could be attributed solely to being older). Despite a long list of problems that can be associated with IVF, at present there is no strong evidence that side effects of IVF are so relevant and/or widespread to restrict its use in healthy couples.

The major clinical problem with IVF is somehow related to its effectiveness, which is multiple pregnancies and births. Multiple pregnancies are associated with greater risks for both mother and foetuses compared with a singleton pregnancy (Edwards, 1998). These include higher rates of caesarean section, low birth weight, and infant death and disability. In the USA in 2003, among the 31,348 pregnancies that resulted from ART cycles using fresh, non-donor eggs or embryos, 59% of pregnancies were singleton, 29% were twin pregnancies, and 6% were triplet or greater pregnancies (Centers for Disease Control, 2003). Thus, overall, about 35% of the pregnancies included more than one foetus.

The high rate of multiple foetus pregnancies results in multiple live births. Thirty-four percent of live births from ART involved more than one infant (31% twins and 3% triplets). This compares with a multiple-infant birth rate of less than 3% in the general US population. The multiple birth rates for IVF are also high in the United Kingdom. In the period 1998/99 about 24% of births were multiple, with almost 3% triplets or more (HFEA, 2006).

There is much discussion in clinical literature about the high rate of multiple pregnancies. The main point of this discussion concerns the positive association between multiple pregnancies and the number of embryos replaced. The HFEA figures for 1998/99 show that the pregnancy rate per cycle with one, two, or three embryos replaced was 9.5%, 26.1% and 26.4%, respectively (table 3). Corresponding multiple pregnancy rates were 3.2%, 26.4% and 33.1%, respectively. Basically, patients and clinicians face a trade-off when they decide how many embryos should be transferred - the higher number of embryos they transfer, the higher the chance of live birth. However, the higher the number of embryos transferred, the higher the risk of there being a multiple pregnancy.

Table 2.3. IVF clinical pregnancy and multiple clinical pregnancy by the number of embryos transferred in the United Kingdom

Embryos transferred	Number of cycles	Number of clinical pregnancies			Clinical pregnancy rate	Multiple clinical pregnancy rate
		Singleton	Twin	Triplet or greater		
One	2,977	276	8	1	9.5%	3.2%
Two	14,144	2,721	959	15	26.1%	26.4%
Three	13,399	2,398	924	248	26.4%	33.1%
Total	30,520	5,395	1,891	264	24.6%	28.7%

Source: HFEA, 2000

In the early nineties UK legislation limited the maximum number of embryos to be transferred to three. Similar rules are enforced in other European countries. Recently, the Human Embryology and Fertilisation Authority lowered the maximum to two. In the USA there is no such limitation thus meaning that almost 50% of cycles involve the transfer of more than three embryos. This may partially explain why the success rate in the USA is higher.

2.6. Ethical and social issues related to IVF

This dissertation focuses on the economic dimension of IVF, using economic analysis to assess whether this technology should be publicly funded. Therefore, a thorough analysis of the ethical and social implications of IVF and other ARTs is beyond the scope of the present study. However, as health policy process concerning reimbursement can be affected by the ethical and social concerns, this section will briefly present the main ethical issues raised by the use of ARTs.

These technologies have made possible to change reproduction in four ways (Koch, 1998). First, technology can be used to substitute, repair or circumvent physical elements of the reproductive process. This is the technology option mainly referred in this dissertation. The technical process is one in which non-donor fresh eggs and sperm are manipulated in vitro with the sole purpose of overcoming infertility.

Second, technology can be used to exchange actors in the reproductive process, making possible the creation of family types that would not otherwise have existed. With IVF using father's spermatozoa and the mother's egg, the child is genetically related to both parents, whereas children conceived with donated sperm are genetically related to the mother but not to the father. Likewise, children conceived using donated eggs are genetically related to the father but not the mother. When both eggs and spermatozoa are donated, the child is not genetically related to either parent. This latter group of children is similar to adopted ones in that they are genetically unrelated to both parents, but differ in that parents experience pregnancy and develop a relationship with the child at the prenatal stage and subsequently birth. Moreover, the exchange of actors in the reproductive process may involve surrogacy: this is when a woman accepts to carry out a pregnancy to deliver a baby that it is immediately given to someone else (the child may be genetically related to neither, one or both parents).

The third class of technical options refers to the possibility to space out the elements of reproductive process by freezing. At present it is technically possible to freeze sperm, eggs and fertilised eggs (embryos). These options make possible to procreate post-mortem. In theory, a baby can be delivered years after the death of both their genetic parents.

The fourth class concerns pre-implantation genetic diagnosis and germ line therapy. Pre-implantation genetic diagnosis is a very early form of prenatal diagnosis aimed at eliminating embryos carrying serious genetic disorders. It can also be used to carry out gender selection. Although not feasible yet, gene therapy may soon allow to fix some genetic diseases at the embryonic stage.

As mentioned above, a systematic discussion of the implications of each of these technical options is beyond the scope of the dissertation. We only briefly present the main issues related to the first typology, which is directly involved by the economic evaluation presented in the following chapters. Before doing this, however, it appears appropriate to make a few general considerations on the overall advancements in human reproduction technology.

In order to neutralise the ethical issues as much as possible, we designed the cost benefit analysis making reference to infertile married couples using their own gametes. Therefore, the economic analysis disregards programmes where donated genetic material is used, that makes use of surrogacy and that include diagnostic activities, embryo selection and cryoconservation. This should have minimised the risk that surveyed individuals were influenced by ethical considerations that are not necessarily involved in conventional IVF (first typology described above). Nevertheless, it should be recognized that such an attempt to counterbalance the ethical issues raised by the other typologies of ARTs cannot be complete. We do expect that some of the respondents were influenced by the overall debate concerning new advancements in human procreation and the ethical issues it raises

The second consideration concerns the relationship between the new advancements in reproduction technology and public funding. The aim of this research is to assess if, from an economic point of view, traditional IVF deserves public funding. Here IVF is seen merely as a way of overcoming infertility. Consequently, we focused on traditional IVF because it is very relevant in economic terms and because it involves few of the ethical problems raised by the other typologies. However, it should be clarified that even more advanced techniques urgently require explicit decision regarding public funding. Let's take for example pre-implantation gene diagnosis and gene therapy. They are extremely controversial as they may take relevant steps towards designing children and eugenics. However, it is also clear that the potential benefits are very relevant as certain diseases can be avoided or somehow made less distressing. In other words, these technologies pose serious ethical problems but also have the potential to provide great health benefits. At present, ARTs are not simply a way of overcoming infertility but also a potential way in which to promote health at the prenatal stage. If the use of ARTs expands as a means to improve people health it will be very important to carefully assess the costs and benefits of public funding. The risk we see in this area is that relevant ethical issues may encourage lesser government involvement in public funding resulting in health inequalities increasing.

Returning to traditional IVF, various ethical issues have been raised. The first, more radical one, concerns the separation of sexual intercourse from procreation. The prime argument against IVF presented by the Catholic Church refers to this separation. According to Donum Vitae (Congregazione per la Fede, 1987), the main Catholic Church document on human procreation, God defined an indissoluble connection between the two meanings of marital life (sexual life): marital union (that is something spiritual) and procreation. As contraception denies the procreative dimension of the marital life and thus it is unethical, in a similar way IVF is unethical too because denies the marital (spiritual) union during the act of procreation. According to the official position of the Catholic Church, sex and procreation are necessarily linked and technology cannot be used to separate them.

This rather radical position appeals mainly to the Catholic community and does not exert sufficient influence at a political level to forbid traditional IVF in virtually all industrialised Christian countries. Indeed, this position was not even endorsed by Italian catholic parties during the parliamentary discussion on IVF in 2004. The Catholic Church, other religions, and also various non-religious cultural and ideological movements appear more rigid with regard to another critical issue related to IVF interventions: the generation and the use of embryos. The in-vitro phase of the procedure aims at producing more than one embryo to facilitate the choice of the best ones to be re-implanted and thus possibly re-implanting more than one embryo and, in some case, cryoconserving them for possible future use. The ethical focus here is on the nature of embryos. Those who consider embryos as a form of life harshly criticise IVF, as in many cases it requires suppressing some of them. In effect, even those who do not recognise embryos as a form of human life generally agree that some form of protection should be granted to embryos and that their handling should be regulated.

Related to the "overproduction" of embryos is also multi-foetal pregnancy reduction. As seen above, in order to increase the chance of obtaining a live birth, very often numerous embryos are transferred to the woman's uterus. This results in a very high incidence of multiple gestations (about 30%) carrying with it increased frequency of complications for the mother as well as a higher perinatal morbidity and mortality. Multi-foetal pregnancy reduction is therefore prescribed to protect the mother, to increase the chance of at least one live birth and sometimes just because the mother is not willing to manage multiple births. To a certain extent multi-foetal pregnancy reduction resembles a type of abortion a can thus criticised and justified on the same ground. However, at least two differences distinguish normal abortion from multi-foetal reduction. The first relates to the origin of the problem. While, "normal" abortion is related to an unwanted pregnancy, multi-foetal reduction derives from a clear desire to have a baby (although it is not necessarily the case that more than one child is desired at a given time). From this point of view multi-foetal reduction may be considered even worse than a "normal" abortion as it contrasts procreation after a series of acts aimed at making it possible. On the other hand, some clinicians claim that multi-foetal reduction should not be considered as an abortion since the purpose of selective termination is the continuation of life, and not the termination of the pregnancy (Fasouliotis and Shenker, 1999). If selective termination is carried out to protect the other foetuses or the mother, it can be argued that the termination is carried out in order to protect life.

The last main ethical issue raised by traditional IVF is the generation of spare embryos that are usually obtained by the method of induced super-ovulation in the woman. The practice of inducing super-ovulation is performed for two main reasons: it avoids subjecting the woman to the trauma and the risks of repeated laparoscopies to recover oocytes (in the case more than one cycle is attempted) and it facilitates the selection of embryos prior to implantation (Shenker, 1997). In addition, the generation of spare embryos may be pursued in order to make spare "material" available for donation or scientific research. Although super-ovulation and thus the generation of spare embryos is not strictly a part of an IVF programme, this is a commonly diffused practice. The generation and use of spare embryos raises the issues of their destiny and their protection and are subjected to a lot of ethical disputes.

2.7. IVF regulation in Italy and other major European countries

IVF and other ARTs are practised in all European countries with the exception of Luxembourg (Shenker, 1997; Spar, 2006). In most countries the practice of IVF commenced in the '80s before specific legislation was enacted. However, major EU members, with the exception of Italy, enacted specific regulation in the late 1980s or in the first half of the 1990s. In France, Germany, Spain and UK specific pieces of legislation authorise and regulate ARTs (Human Fertilisation and Embriology Act, 1990; Lansac, 1996; Beier and Beckman, 1991; Shenker, 1997). In all of the above-mentioned countries FIVET, GIFT, ZIFT and ICSI are legal and can be administered for a therapeutic reason (i.e. to help the couple to conceive if they are infertile). Differences across the various pieces of legislation arise on a number of issues. The United Kingdom and Spain appear more liberal than France and Germany. In the former two countries, in addition to married couples, co-habitant couples and single women can have access to ARTs. In France, access to the techniques is restricted to couples (including those who co-habit), while in Germany assisted reproduction is allowed only to married couples. Legislation in these countries also differs as far as genetic material donation is concerned. The United Kingdom and Spain allow sperm, ovum and also embryo donation; France forbids embryo donation but permits sperm and ovum donation; Germany appears more restrictive as it forbids IVF with donated material. Germany is also more restrictive compared to the other mentioned countries concerning cryopreservation; it is illegal. Whereas, in France, Spain and UK cryopreservation is legal if aimed at making available extra embryos for transfer in a future cycle. In Spain and UK cryopreservation is also allowed for the purposes of pre-embryo research (pre-embryo refers to the first 14 days after fertilization).

In all of the above-mentioned countries selective foetal reduction of multiple pregnancies is practised. However, in order to avoid foetal reduction, all of the four countries' medical guidelines suggest that no more than 3 or 4 embryos are transferred - the UK limits the number of transferred embryos to 2. Another commonality that these countries share is that the national government, regional governments or a specific national authority (like the HFEA in the UK) specifically regulate the activities of the centres that are licensed to provide ART services.

In contrast to the situation of these members of the European Union, Italy has kept infertility services provisionally regulated for a long period. It was only in 2004 that a piece of legislation was passed by the Italian Parliament that related to the regulation of ARTs. The matter was initially regulated by a Circular enacted by the Ministry of Health that was conceived as a provisional measure, further it was targeted only to the NHS. However, due to the lack of legislation, the Circular became the judicial reference for the practice of ARTs in the NHS and the private sector was left without regulation. As a result, until 2004, the practice of IVF in the NHS was strictly limited, while the private sector could act well beyond what is legal in most of the other EU member states.

The Circular, while introducing the principle that infertility treatment is part of the duties of the NHS (thus implying that public coverage was assured), clearly limited the availability of the treatment to married couples and excluded the use of donated material. At the same time, the lack of explicit regulation on the matter in the private sector left ample room for the establishment of private practices that performed an array of services, including services to unmarried couples and single women, IVF and other ARTs with donated material, cryopreservation, and even gender selection.

Although there is insufficient evidence is, it appears likely that Italy was not a safe environment for couples (and individuals) seeking infertility treatments. The lack of legislation and the strict rules imposed, limited the provision of IVF and other ARTs in the public sector, whilst paradoxically it promoted the establishment of private clinics that could operate with minimal surveillance and control. Although professional associations established codes of practice (CECOS, EFRA Italia), they were not mandatory and many private practices operated at their discretion, with no obligation to respect standards and no explicit accountability systems. Despite a National registry of ART centres being established by a government directive, registration was not made compulsory therefore there were many centres that were not even known by the Ministry of Health and Regional Health Authorities.

This situation has been highly criticised by professional associations, patients' advocates and all political parties. However, for many years the Parliament was not been able to pass legislation, despite the fact that many members of Parliament had officially stated that legislation was absolutely required to stop the "Italian Wild West" of reproductive services.

Indeed, in 1999 it looked as though legislation would have been enacted. The "Camera dei Deputati" (the Lower House) approved an act covering all the main issues and in the "Senato" (the Upper House) major political parties stated that they would have voted in favour of this piece of legislation as it was the result of a reasonable compromise between different instances. Instead however, members of the Senate claimed autonomy from the official party positions on the grounds that, on such vital issues as those raised by ARTs - personal value judgements were more important than party discipline. Consequently, the equilibrium reached in the Lower House dissolved and thus the proposed legislation was not approved.

Legislation was finally passed in 2004, after fierce parliamentary discussion (Legge 40, 2004). It clearly states that Assisted Reproduction Technologies are allowed only to overcome fertility problems and only once other therapeutic methods (e.g. artificial insemination) have failed. It permits IVF and other ARTs to married and stable heterosexual couples, but only using the couples' genetic material. Donation of eggs, sperm or other genetic material is not allowed and severely punished. The legislation also bans the freezing of embryos, limits the number of embryos that can be implanted in the woman's womb to three and forbids embryo research. Opponents to the legislation tried to cancel it through a national referendum held in June 2005. The referendum failed because only 34% of adult Italians (automatically registered to vote) went to the polls. Basically, those who favoured the approved legislation campaigned against going to the polls. As a consequence, despite a sweeping majority among voters, abolishers did not reach the *quorum* of the majority of Italians registered to vote and hence the legislation was not abolished.

2.8. Public funding of IVF

In the past IVF was often used to discuss criteria for rationing (Klein et al. 1996, Giacomini et al., 2000). The Dunning committee in the Netherlands suggested that IVF was not necessary and thus should be excluded from public coverage because childlessness does not interfere with normal functioning in the Dutch society (Van de Ven, 1995). President Clinton's ill-fated Health Security Act specifically excluded IVF services from the minimum package of services to be covered by insurance plans (Spar, 2006). In the early nineties, in a UK study of 114 purchasing plans, the majority of Health Authorities stated that they were purchasing some IVF (Redmayne et al., 2003). However, six health authorities explicitly stated they did not want to buy IVF or other similar ARTs. This old study clearly showed that the public funding of IVF was not consistent across the UK. More recently, a Canadian study used the case of IVF to study the multiplicity of meanings of not insuring services and to show the complexity of rationing decisions.

In the early days of ARTs many governments tried to exclude them from public coverage on several different grounds. These included the claim that infertility is not a disease, that IVF is not a cure but only a way to bypass infertility problems, that IVF is not really needed because some people decide to be childless, or that, in those days, IVF was still experimental. Indeed, IVF was an interesting object of rationing because it is peculiar for the "need" it addresses and for the outcome it promises. Obviously, procreating babies is different from saving lives or improving the quality of life of people suffering because of a disease. It is not denied that people with infertility may live in stressful conditions, but it is argued that it is purely subjective as some people prefer not to have babies and are happy about that. Generally, diseases do not work this way. We could not find other examples in which people are not negatively affected by having a pathological condition and may be even pleased by that. Also, ARTs interfere with procreation rather than with existing lives and this, as seen earlier, may pose ethical problems. Whatever the reason, IVF was a preferred target for discussion and for explicit decisions about rationing.

At the moment, IVF is provided, although with limitations, with some government or social insurance funds in the four largest EU member states (France, Germany, Italy and the UK). None of these countries entirely excludes IVF from public or statutory coverage, even if different rules may apply about eligibility criteria (e.g. the age of the woman), types of ARTs offered and maximum number of cycles per patient. Nevertheless, it appears likely that publicly funded supply falls short of demand. There is some evidence to suggest a lack of access to treatment within the British and the Italian NHS, the two largest tax-funded systems in Europe. The HFEA (2006) reports that about 25% of IVF treatments are funded by the NHS. This low percentage contrasts with NICE recommendations that couples should be offered up to three cycles of IVF on the NHS if the woman is aged 23-39 and the couple has an identified cause for their infertility, or have not conceived after three years. There are no reliable data in Italy about the use if IVF, as at June 2006 a list of all the centres nationwide that the Italian regions have authorised to provide IVF treatments was not available. Nevertheless, three clinicians that we met when collecting cost data reported that the number of NHS centres is rather limited and that the majority of treatments are performed by private institutions and are paid for directly by patients.

It appears that IVF treatments are not denied by official government positions in major European countries, but rationing may take place in the form of short supply and lack of compliance with the recommendations. In this respect, NICE in England and Wales took a clear stand, as it made specific recommendations about the use of IVF treatments, while the Italian NHS appears generic in the way it includes them in its basic package. The following chapters will present and discuss methods and results of a cost-benefit study designed to offer guidance on the issue of public coverage of this treatment. From an economic perspective it is based on the assumption that a thorough computation of costs and benefits according to a welfare economics model may allow us to discern whether the programme makes a positive contribution to the welfare of a defined population. While we recognise that there may be merits in some forms of uncertainty and ambiguity in NHS coverage (Mechanic, 1995), we strongly believe that systems need to be more "rational" in the way they use scarce resources. We agree with New and Le Grand (1996) that there is a real danger that moving too fast towards explicit, systematic and democratic forms of rationing may be problematic because the "NHS is about more than simply producing health ...(as) it acts as a mechanism, and a symbol, of reassurance and social stability to that sound economic analyses may help decision making". We have not designed and conducted this study with the intent to offer policy-makers the final solution about IVF coverage. We simply think that a sound economic analysis, designed to help decision-making, may be a relevant input to make better choices in health care systems.

Chapter 3

The contingent valuation method for measuring the benefits of health care

3.1. Introduction

Since the seminal work by Mishan published in 1971 applications of cost-benefit-analysis (CBA) have been mainly based on the estimation of individuals' willingness-to-pay (WTP) or willingness-to-accept (WTA). The former refers to the maximum amount of money that individuals are willing to pay for a set of benefits, while the latter refers to the minimum amount of money that individuals are willing to accept as compensation for a set of losses. There are two main empirical approaches that can be used to measure WTP or WTA for benefits attributable to healthcare programmes: revealed preferences and expressed preferences. Revealed preference studies indirectly measure WTP or WTA through the observation of actual behaviour. Choices concerning occupation, housing, diet and driving behaviour can reveal people's attitude towards their health and, hence, can provide data to estimate WTP or WTA for some health hazards. The most common type of revealed preference studies related to health refers to wage-risk trade-offs on the labour market (Johannesson 1996a; Viscusi 1993).

Expressed preferences studies derive WTP or WTA from survey questions referring to hypothetical markets. This approach circumvents the absence of markets for some goods and can capture elements that are not present in private market decisions. Surveys can be designed to reflect either a private goods market or a political market (Mitchell and Carson 1989; Smith, 2006 see below). Since the values elicited from these surveys are contingent upon the particular hypothetical market described to the respondent, this approach has been called the Contingent Valuation (CV) method.

The first trace of the idea to elicit monetary benefits from hypothetical surveys and to use them in cost-benefit analysis dates back to 1947, when Ciriacy-Wantrup wrote that:

"Individuals of a sample or of a social group as a whole may be asked how much money they are willing to pay for successive additional quantities of a collective extra-market good. ...... If every individual of the whole social group is interrogated, all individual values (not quantities) are added. The results correspond to a market-demand schedule... In combination with a corresponding cost schedule the socially desirable supply of the collective extra-market good can be determined."

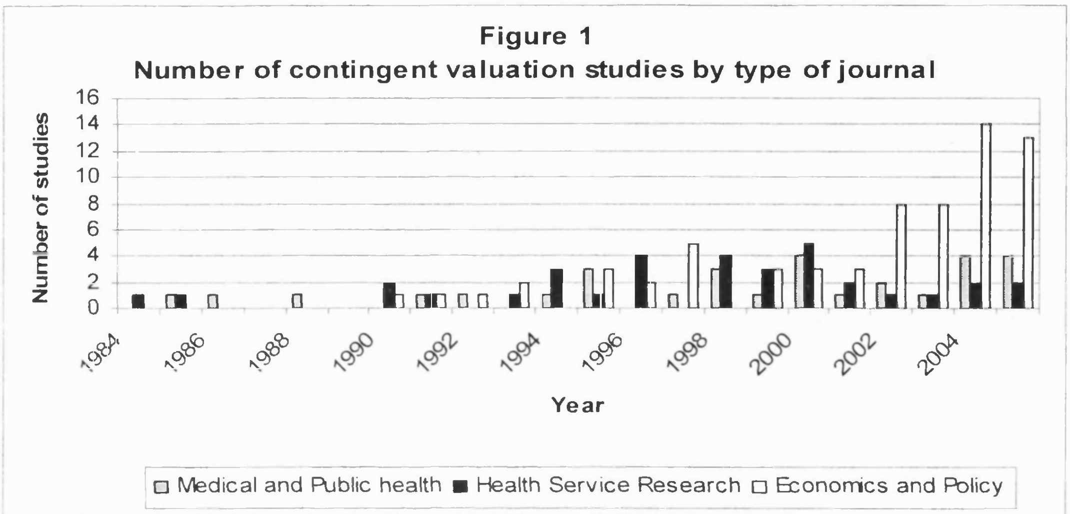
Starting from the mid 1960s the CV method has been developed both at the conceptual and empirical levels (Smith, 2006). The development of the method has mainly taken place in two areas: transport economics and, more recently, environment economics. In transport economics the main applications of the CV method have focused on measuring safety benefits and the value of travelling time. In environment economics, instead, studies have focused on the non-user values of environmental habitat and resources and, to a lesser extent, health benefits stemming from improved quality of the environment.

Davis (1963) is the first person who used the method. He made a contingent valuation survey in Maine in order to measure the value of a recreational area. Since then, several hundreds of papers have applied the Ciriacy-Wantrup intuition in the environment field. The first contingent valuation survey related to the provision of health care was performed by Acton (1973) in the United States. The analysis focused on a programme making available mobile coronary care units that would reduce the risk of death after a heart attack. Acton designed a survey with various types of WTP questions and administered it to three different samples: one that was a randomly stratified sample of households living in the Boston area (36 households) and the others that comprised union leaders (21) and business executives (160). The response rate was very low with only 36 individuals returning the questionnaire. Moreover, 11 questionnaires reported a willingness-to-pay of zero and were considered by Acton as protest answers. On the basis of these and other elements Acton concluded that the validity of the study was unclear.

Either because of the problems of Acton's study or because researchers were more attracted by the development of cost-effectiveness and cost-utility analysis, for many years no other studies used the CV method for valuing health care. The second attempt to use the method was made by other Harvard scholars ten years later:

Thompson, Read and Liang (1984). They investigated WTP for a hypothetical cure of chronic arthritis. Since then the CV method had not been substantially developed in the health sector for many years. It is only in the last 10/15 years that papers using contingent valuation techniques have become more frequent in health economics, public health and clinical journals (figure 3.1).

Figure 3.1. Number of contingent valuation studies by year of publication and type of journal



In fact, the number of contingent valuation studies in health care is growing. This is probably due to a number of factors, including the evident limitations of cost-effectiveness and cost-utility analysis, the consistency of the CV methods with welfare economics theory, and the improvement of the techniques of surveying people and of sampling individuals. In addition to these reasons, one merit of CV methods is the possibility to investigate directly the preferences of members of the community. These surveys mimic procedures of hypothecation of government spending. If hypothecation is desirable because it turns pawns into queens (or at least it goes some way towards doing so) (Le Grand, 2003), eliciting preferences for specific health programmes is expected to offer guidance to policy makers making them aligned to individual preferences.

Contingent valuation is now performed by economists and other specialists in different areas, and it is used by government agencies and international organisations for the economic evaluation of a variety of investments. Carson's (1994) bibliography listed about 1,000 studies over 40 countries in several fields, including the environment, transportation, the art and education, sanitation and health. However, until the end of the '90s very few contingent valuation studies concerned health services. By the means of a search strategy based on the two most popular computerised databases in medicine (Medline) and economics (Econlit), and bibliographies of some of the retrieved studies, we collected only 33 papers published from 1984 to 1996 which report results of empirical contingent valuation studies. Nowadays, the production of CV studies is more significant with about 30 papers published annually.

This chapter reviews studies that have directly elicited WTP or WTA for healthcare services through the CV method. Over the past ten years it has become increasingly popular in the health economics literature to elicit WTP through an indirect method labelled Discrete Choice Experiment (DCE). DCE belongs to the family of conjoint analysis (CA) techniques. These techniques have their origin in market research and are used to establish the relative importance of different attributes in the provision of a good (Ryan, 1999). They have been used by non-economists to investigate factors that explain patients' preferences for attributes of health services and by economists to derive the monetary value of benefits associated to the use of health services and, more generally, to measure preferences. When CA is used within the framework of economic theory it is generally labelled DCE or sometimes CE (Choice Experiment). DCEs have been used in health economics to estimate WTP to undergo IVF treatments (Ryan, 1999), to reduce time spent on waiting lists (Propper, 1990), to investigate preferences and to test WTA measures for blood transfusion services (van der Pol and Cairns, 1998), to elicit patient preferences in the doctor-patient relationship (Vick and Scott, 1998), to estimate time preferences for health (van der Pol and Cairns, 2001; Cairns and van der Pol, 2004), and as a method of deriving preference-based values for process of care and health state outcomes (McKenzie et al., 2001, Hakim et al., 1999).

This chapter focuses in CV studies only and reviews the literature concerning health services. In addition, it selectively reviews the environmental econonomics literature on three specific issues: the possible formats to elicit WTP, the use of internet to conduct CV studies and the assumptions about the probability distribution functions of WTP in case of zero WTP. In these specific areas the CV literature in the environmental field could provide us with relevant material to understand the nature of the data that we collected and to identify the appropriate econometric models. We also checked other fields such as the art and transports, but we realised that developments in contingent valuations have taken mainly place in environmental studies.

3.2. A framework to review contingent valuation studies

There appears to be large methodological differences in the health care contingent valuation literature on important issues such as the content of the WTP questions, the way they are asked and the choice of the sample (O'Brien and Gafni, 1996). This is because the CV method is a young research field and thus it is very far from being well established from a theoretical point of view. Nevertheless, in order to carry out a literature review of contingent valuation studies in the health care field, it is necessary to have a classification framework from which to start.

The growing number of empirical studies in the health care field has been critically discussed by a few methodological reviews (Diener et al., 1998, Klose, 1999; Olsen and Smith, 2001; Smith, 2003). Some of these contributions are discussed later in this chapter. In this study, the framework proposed by O'Brien and Gafni (1996) is used to classify and review the contingent valuation literature in health care. This critical appraisal framework presents eleven specific considerations grouped in five general questions aiming to help in the interpretation of the design of the contingent valuation studies. This framework does not discuss the advantages or disadvantages using the method, nor does it address the theoretical foundations of CBA because it accepts that as given. It presents some basic questions whose answers may help to understand the main conceptual and methodological issues related to the use of contingent valuation method in the health care field. In addition to this contribution, the analysis of the literature is based on an important contribution specific to the environment field: the document prepared by a panel of economic experts to evaluate the use of the CV method in determining non-use values of environmental resources and provide comment to the National Oceanic and Atmospheric Administration (NOAA) of the USA Federal Government (Arrow et al., 1993).

The conceptual framework suggested by O'Brien and Gafni (1996) (Diener et. al, 1998) is structured around eleven considerations grouped into five general questions (table 3.1):

1) What questions do we want to answer?

2) What type of measure can we use?

3) What do we need to ask of whom?

4) What characteristics of the programme are important to determine how it is valued?

5) What question formats minimise bias and increase precision?

Questions and considerations are presented and discussed below according to the original paper.

Table 3.1. Questions and considerations for a contingent valuation study of a health care programme (O'Brien and Gafni, 1996).

1. What question do we want to answer?	Problem definition Pricing and demand studies Project appraisal for resource allocation Current status of the programme Programme currently exists Programme does not currently exist Utility and disutility of programme to respondent Gain in utility from programme Loss in utility from programme
2. What type of measure can we use?	Money measure of utility change Compensating Variation (CV) Equivalent Variation (EV) "Direction" of measurement Willingness-to-pay (WTP) Willingness-to-accept (WTA)
3. What do we need to ask of whom?	Framing of programme consumption and payment Ex-post user-based question Ex-ante insurance-based question
4. What characteristics of the programme are important for determining how it is valued?	Programme outcome description Certain outcomes Uncertain outcomes Nature of the "market" for valuation scenario Private goods market Political market
5. What question formats minimize bias and increase precision	Valuation scenario Holistic versus decomposed Degree of realism Value elicitation method Open-ended questions Bidding games Payment cards Take-it-or-leave-it (with follow-up)

3.2.1. What question do we want to answer?

Problem definition.

WTP may be studied for a variety of reasons and from a number of different perspectives. Not all studies trying to elicit WTP for health services are carried out as part of a cost-benefit-analysis. Some monetary valuation studies are carried out to estimate private demand curves and, hence, to support a variety of policy decisions including the use of co-payments. More broadly, there are some studies that are concerned with marketing decisions rather than with collective decision-making about health care programmes.

Although there are some WTP questions that may be unrelated to economic evaluation, we found it conceptually difficult to disntiguish whether the underlying question is one of marketing or evaluation. Marketing is more than pricing or promotion; it involves studying the expectations and other characteristics of consumers. As economic evaluation techniques are applicable to any collectively-funded delivery system (Pauly, 1995), the possibility that private insurance companies use CBA or other economic evaluation techniques to offer guidance to the definition of insured risks and services should not be excluded. Also, strategic marketing decisions in the pharmaceutical industry and other industries related to healthcare can be supported by economic evaluations according to different perspectives. In addition, marketing is developing in non-profit organisations as well in the public sector. Organisations whose main aim is not the generation of profit are increasingly putting more effort into better understand their "consumers" and to use "market" information to support their decision-making. As a consequence, in public and non-profit organisations economic evaluation can be considered as a "social marketing" technique.

Many published papers are focused on methodological issues such as the validity of different elicitation formats or the characteristics of the scenarios presented. These studies deserve to be considered separately from the others. Therefore, for the purpose of the literature review presented in this chapter, we have reformulated the classification suggested by Gafni and O'Brien (1996, Diener et al. 1998) on the problem definition of the studies. We prefer to classify studies according to four categories: a) economic evaluation (the contingent valuation method is used as part of a cost-benefit analysis); b) analysis of demand (the method is used to investigate the characteristics of the demand for the programme); c) cost-of-ilness studies (the analysis is used to estimate the economic burden of the disease, d) methodological studies (the study is focused on testing the contingent valuation method).

Current status of programme

Some studies regard programmes that currently exist, others refer to programmes that are hypothetical. The utility gain of a project perceived by the respondents of the CV survey may be different in the two scenarios. It is thus useful to distinguish the studies according to the following classification: a) existing programmes (the object of the evaluation is available or could be available to the respondents); b) hypothetical programmes (the survey presents a programme that is hypothetical).

Utility and disutility of the programme to respondents

It was argued that in the case of health care, negative bids can be excluded (Johannesson and Jonsson, 1991). However, not all individuals will be affected or will be better off from the introduction of a new health care programme. Similarly, the cessation of an existing programme may be a source of utility for some people. The removal of public coverage of voluntary abortion would represent a utility gain for some people. Also, some programmes prolonging the life of individuals with infectious diseases bring disutility to many members of the society that may face an increased risk of infection (Villari et al, 1996). With particular reference to our study, it is possible that some reproductive techniques face ethical opposition and are thus associated with utility losses (for example the introduction of a new programme) or gains (in the case of removal).

It is therefore important to check whether the decision under evaluation brings about losses or gains. In a situation where they are both present, appropriate devices have to be used to elicit the monetary equivalent of both positive and negative utility changes. Unfortunately, it proved very difficult to derive information in this respect from the paper because this issue has been neglected by many authors, even where it was evident that potential losers were likely. Therefore, we decided not to include this issue in the systematic analysis of the health papers and to discuss it in the final part of the chapter.

3.2.2. What type of measure can we use?

Monetary measure of utility change

Theoretically, it is possible to measure willingness-to-pay (WTP) or willingness-to-accept (WTA) on the basis of two monetary measures of utility change: *compensating variation and equivalent variation*. The distinction depends on whether the utility level is held constant before or after the decision is implemented. If the analysis is based on the starting utility level (the utility level before the change), WTP and WTA measures are usually referred to as compensating variations. On the contrary, if WTP and WTA are measured according to the new level of utility (the level of the utility after the change), the monetary measures are usually called equivalent variations. It should be noted that, although they are equivalent measures of consumer surplus if income effects are irrelevant, they do differ in perspective: compensating variation is evaluated from the original level of welfare and equivalent variation from the anticipated new level of welfare (O'Brien and Gafni, 1996; Diener et al., 1998).

Direction of measurement

The distinction between equivalent variation and compensating variation is further complicated by the distinction between WTP and WTA. As mentioned above, WTP is the maximum amount of money that individuals are willing to pay for a set of benefits, while WTA is the minimum amount of money that individuals are willing to accept as compensation for a set of losses. WTP questions are generally framed as "How much would you pay to have this set of benefits?", while WTA questions are framed as "How much would you accept as compensation for giving up these benefits?" The two measures are not equivalent if the marginal utility of income is assumed declining. As WTP and WTA are utility measures they are affected by the direction of change in income (Brent, 2003). If the utility of income is decreasing, for a given good WTP is greater than WTA because the consumer is available to sacrify more income to get the good that she is available to receive as compensation. If the marginal utility of income were constant, then there would be no difference between WTP and WTA.

3.2.3. What do we need to ask of whom?

Externality and option value

The framework suggested by O'Brien and Gafni (1996) comprises three types of individuals whose WTP or WTA can be measured: individuals who have the disease; individuals who have not the disease, but they are at risk of acquiring it; individuals who have not got the disease and who are not at risk of acquiring it. The first group of individuals are the immediate beneficiaries of the health care programmes. They are expected to express a strong preference for programmes which are in their favour. Accordingly, they are generally expected to state a higher WTP than people without the disease. The second group of individuals refer to the option value concept (Weisbrod, 1964). People may be willing to pay an amount to ensure that the health care programme will be available at a later time. This option increases people welfare because, by acting like an insurance policy, it reduces uncertainty. The third type refers to individuals who are not and will not be affected by the disease but are somehow affected by its consequences. Externaility is the economic concept involved here (Culyer, 1976). Two types of externalities can be distinguished: physical externalities and altruistic externalities. The first type of externality refers to situations where some individuals are physically affected by the programme even if they are not the direct beneficiaries. Examples of programmes bringing about such externalities include prevention programmes for infectious diseases targeted to specific population groups (the community at large benefits as a result of lower incidence rates), wide and systematic use of antibiotics (the community at large develops resistance to the drug), treatment of individuals affected by certain mental disorders (keeping these individuals on treatment can decrease the risk of criminal offences). This type of externality is only concerned with the individual's well-being.

The second concept of externality assumes a broader perspective - that individuals can derive utility from the fact that other individuals have access to specific goods. Altruistic externalities are probably very relevant in health care since public funding is often motivated on the basis of solidarity principles. In addition to positive altruistic externalities, it is conceivable that people may suffer from the others' consumption of specific goods which raise serious ethical problems, such as voluntary abortion, IVF or even organ donation (see later in the chapter).

It should be clear that these three "groups" of beneficiaries reflect three components of benefits of health care programmes that may co-exist in the same individual. A patient affected by a disease may reveal herself as being willing to pay for the care she needs (patient value), thus ensuring that the option is available for her in the case of future personal need (option value) and for making it available to other people in state of need (altruistic value).

As far as the classification of the paper is concerned it proved difficult to detect whether the option value component was present. Thus, we simplified the classification originally proposed by O'Brien and Gafni (1996) focussing on whether there was any altruistic component on the measuered WTP/WTA.

Framing of programme consumption and payment

There are two general approaches that can be used in a CV study: the ex-ante insurance-based and the ex-post user-based. In the ex-post approach, respondents are asked to assume that they are at the point of using the service being evaluated. Under this hypothetical scenario they are asked to state their WTP or WTA. In the second approach, respondents are asked to express their WTP or WTA for an insurance that covers the service being evaluated. Obviously, the second approach requires providing the person with relevant information about the probability that the service will be used.

In an earlier paper, Gafni (1991) suggested that in the context of public decision-making the ex-ante insurance based approach has to be followed. This suggestion stems from the consideration that users of health services do not generally pay at the point of consumption for the services they use. Also, by definition public funding of goods whose future utilisation is uncertain is a form of insurance. In spite of this position, however, the conceptual framework presented here recognises that there are some circumstances where asking insurance-based questions may not be appropriate. For example, most individual and collective prevention programmes cannot be insured because their consumption is not influenced by any uncertain events. However, for all the services that can be subjected to an insurance service, O'Brien and Gafni (1996) suggest using the ex-ante perspective because the "*user* *approach is not consistent with the theory CBA, which is based on the sum of the compensating (or equivalent) variations from the population affected*".

The ex-ante approach is also preferable because it allows the detection of respondents' altruism. However, there are relevant problems with eliciting WTP from individuals who are not experiencing the disease: probabilities are difficult to understand by most people and the general population has less information about a specific disease and treatment than a member of the patient population. One way to overcome these problems is to use an ex-post user-based perspective and to multiply elicited WTP (or WTA) by the probability of becoming a patient. Following this procedure, the expected WTP indirectly measures non-user value (but it does not detect any externalities and does not include any valuation of the "option value").

As shown by Johannesson (1996b) if the individual is risk neutral, ex-ante WTP and expected WTP coincide; while if the individual is risk averse with respect to income, expected WTP is a lower bound of ex-ante WTP, provided that the marginal utility of income does not vary with health status. These results should be taken into account when empirical studies are interpreted. Under the assumption that individuals are risk averse, if costs are lower than expected WTP the programme also passes the cost-benefit test from an insurance perspective.

Unfortunately, it cannot be assumed that expected WTP is a lower bound. First, as mentioned above, questions directed to users do not capture externalities, as in some cases the sign of the externalities can be uncertain. Second and more important, the assumption that the marginal utility of income does not vary with changes in the health status appears strong. Our knowledge about the interaction between marginal utility of income and health status is limited (Johannesson, 1991). A study conducted by Viscusi and Evans (1990) shows that the marginal utility of income increases with better health status. However, it may be the case, especially for certain diseases, that individuals perceive income as a substitute for health. Thus, health improvements are associated with lower marginal utility of income. Given that health is a very broad concept, often involving psychological aspects specific to the type of disease, it is probably difficult to identify general regularities between health status and marginal utility of income. Consequently, the possibility that WTP is a lower bound for ex ante WTP, even if externalities are ruled out, should be investigated in the specific context of the programme being evaluated.

While the insurance perspective may be theoretically relevant, in practice very few studies are framed according to this perspective. It appears to be important to distinguish between patients with and without experience of the intervention being evaluated. Hence, we classified studies according to three categories: ex-ante user based (the respondent has not experienced the intervention yet, but she is in need of it), ex-ante insurance based (the question refers to paying for insurance for the intervention), ex-post insurance based (the respondents were interviewed after they had experienced the intervention being evaluated) (Diener et al., 1998).

3.2.4. What characteristics of the programme are important for determining how it is valued?

Programme-outcome description

Under this consideration a distinction is proposed between certain and uncertain outcomes presented in the survey. The distinction was suggested by Gafni (1991) who argued that contributions to health (both positive and negative) attributable to the programme should be described in probability terms. However, it has to be taken into account that it may be very difficult to present uncertain outcomes, especially in ex-ante surveys were the presentation of the uncertainty of the outcome is added to the information concerning the probability of acquiring the condition for which the treatment is used.

Adequate presentation of the uncertainty surrounding events is one of the elements which have to be carefully thought about in designing the CV study. Other important aspects to be carefully described in the programme-outcome presentation include time and space specification, available alternatives to the programme, and a precise description of how the programme is funded.

Nature of the "market" for the valuation scenario

Scenarios used in contingent valuation studies should be consistent with the nature of the decision being evaluated. Guidelines with regard to the use of contingent valuation for the measurement of environmental damage suggest adopting a political market perspective: CV scenarios should be built in such a way that the respondent is asked whether he/she accepts to be taxed a certain amount (or a percentage of income) to fund a specific project (if WTP is measured) (Arrow et al, 1993). The decision to refer to political markets is mainly due to the technical nature of environmental goods. Being non-rival and non-excludable, they are subjected to free-riding in the private market, hence meaning that they require government intervention to be adequately funded.

For certain health services the same argument applies. Most of health services, however, are private goods in nature because they do not conform to non-rivalry and non-excludability criteria. Therefore, for health services, it is generally possible to design CV surveys that describe private market scenarios. These scenarios mimic a private market asking respondent whether he/she is willing to pay (or accept) to have access to a specific good (including insurance). Indeed, these scenarios are probably preferable if the aim of the study is to measure *user* WTP or WTA. But, as noted earlier, often economic evaluations assume a public decision-making perspective and, consequently, aim to measure externalities, both physical and altruistic. In these economic evaluations the use of private market scenarios would not allow for the capture of the altruistic value of programmes being evaluated.

3.2.5. What question formats minimise bias and increase precision?

The survey instrument in a CV study is of the utmost importance (Johannesson and Jonsson, 1991). There are a number of sources of potential bias that are important (Mitchell and Carson, 1989). O'Brien and Gafni (1996) focus on two aspects that involve tradeoffs between precision and bias.

Valuation Scenario

The scenario presented in the survey can be holistic or decomposed. The former type of scenario needs to present the entire complexity of the programme. Consequently, WTP (or WTA) elicitation derives from a very complex process of evaluation which is largely unknown to the authors of the study. For complicated programmes this process can be very demanding from a cognitive point of view. On the contrary, in decomposed scenarios, the components of the programme are evaluated separately so that respondents may be presented simpler questions.

There is no algorithm for aggregating decomposed WTP. In fact, they report one of the author's experiences in carrying out a CV survey to measure benefits of a new antidepressant therapy (O'Brien et al, 1995). A decomposed scenario was presented to respondents because in the pilot study the holistic approach was found to be too cognitively demanding. However, despite the fact that the questions were easier to understand, the result of the survey was inconclusive on the value of the new drug because it is unknown how to aggregate single elicited WTP. Empirical studies generally show that the holistic WTP is sub-additive: adding up decomposed WTPs overestimates elicited values (Klose, 1999; Hammerschmidt et al., 2004).

In our opinion, decomposing scenarios is rarely a viable option in the CV methodology. The fact that some scenarios can be too cognitively demanding does not derive from the hypothetical nature of CV, but rather by the fact that complexity is a normal characteristic of real markets. Real purchasing or referendum decisions are often cognitively very demanding. In spite of this, decisions are taken and economic theory assumes that they are rational. In order to deal with cognitively demanding scenarios it is thus preferable to make the hypothetical markets closer to reality, rather than simpler but imaginative. Accordingly, it is suggested that the way the scenario is presented should be as realistic as possible. In particular it is suggested that scenarios should make sense in the context of the health care system and method of funding with which the respondent is familiar.

Whether the scenario is holistic or decomposed and the degree of realism are important characteristics of the CV survey. In addition, other elements are important to improve its validity and reliability.

Valuation-elicitation method

Gafni and O'Brien (1996) described five elicitation methods used in CV surveys: open-ended questions, bidding games, payment cards, and take-it-or-leave-it questions with or without follow-up. Recently, there have been attempts to use payment ladders (Hanley et al., 2003) and it has proposed a new elicitation method termed "structured haggling" that is assumed to better reflect the bargaining nature of purchasing decisions in the context of a sub-Saharan country (Onwujeckwe, 2004).

Open-Ended Valuation Questions

In open-ended (OE) valuation questions each respondent is simply asked his/her maximum WTP for the programme. The approach is efficient, in the sense that it produces a large amount of information with simple and easy to administer questions. However, open-ended questions pose a large cognitive task and produce large number of non-responses or protest zero responses. It was described by the NOAA panel as "ulikely to provide the most reliable valuations" because it poses an exptremely difficult task, is unrealistic and is prone to strategic overstatements. In effect, in the real world goods are priced and consumers rarely formulate in their mind a maximum willingness-to-pay amount for goods. The OE format has been largely dismissed after the publication of the NOAA guidelines. However, its inferiority to other formats has been contested (e.g. Ready et al., 1996).

Bidding Games

The second elicitation method -bidding games- resembles auctions. An initial starting money bid is made to the respondent who accepts or rejects it. The bid is then raised or lowered depending on the answer. If the answer is "yes" the bid is raised and if the answer is "no" the bid is lowered. Bids are presented to the respondent until his/her willingness-to-pay is reached, that is when a lower/bound is reached or when there is a switch (from "yes" to "no" or viceversa). The method is more realistic than the open-ended format because each bid level requires only a yes/no response. Its main disadvantage is the potential for a starting bias, as the initial bid tends to imply a value for the good. One solution may be to vary initial bids across the sample or to let the respondent choosing the starting value (Bennet et al., 1996; Bennet & Trantert, 1998).

Payment Card

The payment card method visually presents to the respondent a range of values to choose from. This method can be conceived as a facilitated open-ended format. Respondents have to decide a value (or a range of values), but these are somehow presented by the questionnaire or the interviewer. In this way the potential for a starting bias is reduced. Nonetheless, the unrealistic nature of the open-ended question remains.

Take-it-or-leave-it (dichotomous choice or referendum)

In the take-it-or-leave-it (TIOLI) approach each respondent is asked whether he or she is willing to accept a single bid. By varying the bid in different sub-samples, the proportion of respondents who are willing to pay different bids (prices) can be calculated so to identify a relationship between bids and fractions of the sample. This curve can be interpreted as an aggregate demand curve (Johannesson, 1996a). The TIOLI approach is also named dichotomous choice or referendum format. There are three main problems with this approach. First, it is inefficient because, compared to other elicitation methods - a much larger sample size for the same level of statistical precision is required. Second, it requires important assumptions about how to specify the demand curve (Mitchell and Carson 1989; Donaldson et al., 1998). Third, it may result in an inflation of mean WTP amounts as respondents may favour "yes answers" simply to please the interviewer or because they feel good about revealing their willingness-to-pay (warm-glow effect) (Desvousges et al., 1993; Hanley et al., 2003; Ryan et al., 2004). Nevertheless, this elicitation method appears to be close to actual behaviour (we generally decide whether to buy at given prices) and is very easy to understand by respondents. The NOAA panel for the evaluation of environmental damages suggested adopting this approach (Arrow et al., 1993) and is still widely used.

The take-it-or-leave-it question can be followed by one or more bids. The respondent is asked a question requiring a "yes" or "no" answer about whether she/he would pay a specific price. If the respondent states "yes", another WTP question is asked using a higher price chosen. If the answer is "no", the follow-up question presents a lower price. By collecting more answers from each respondent, the method improves statistical efficiency of the simple binary approach and resembles a bidding game.

More recently, there has been an interest in using payment ladders (Hanley et al, 2003). Such ladders suggest a range of uncertainty over the value respondents express on the commodity being valued. The rationale for using this elicitation method is based on the assumption that while respondents may know for sure whether they would or would not be willing to pay a given amount for a good, there are other amounts that they would not be sure about. For instance, a respondent may be sure that she would pay 20 and that she would not pay 100, but she could be unsure about whether she would pay 50. This method although little used in evaluating health care has used and tested in environmental studies and is often lable "multiple bounded" (Welsh and Poe, 1998; Cameron et al., 2002).

The Structured Haggling Technique (SH) was developed by Onwujekwe (2004) with specific reference to the sub-Saharan context. Price-taking in many African and other developing countries differs considerably from the context in which binary questions or payment cards have been suggested. Question format is a major element of contingent valuation surveys which is unlikely to be transferred successfully without paying attention to the environment circumstances. Onwujekwe (2004) argues that the elicitation format needs to be indigenous to the area in which it is used. According to the proposed SH technique, the interviewer initiates the haggling process by offering the good to the respondent at a price that is well above the expected sale price. Then the method tries to simulate a real bargaining process that is expected to terminate with a value that reflects respondent's maximum willingness to pay. This method appears closer to the way markets function in many countries; it also has the merit to guide the respondent in constructing his/her WTP. Psychologists, behavioural theorists and recently economists have argued that preferences are not generally "ready" to be discovered; on the contrary, they are constructed by individuals in response to stimuli (Cookson, 2000; Ryan et al., 2004).

3.3. A literature review of contingent valuation studies in the health care field

3.3.1. General information on the search strategy and the collected studies

We reviewed the literature following three main strategies. First, we collected empirical articles published in major journals, methodological contributions and reviews to understand the debate upon the CV method and its critical issues. This part of the literature review aimed to provide the background for the empirical study presented in the thesis. Then, we systematically collected studies published from 1984 to 2005. This review allowed us to investigate how CV empirical studies have changed over time and to implement the framework presented in the previous section. Finally, we reviewed in depth all CV studies that deal with IVF, the focus of our empirical research.

We performed the literature review in two different moments and we used two different research strategies. We collected the contingent valuation literature for the 1984-1996 in 1998 according to three different strategies. First, the English-language literature was searched in two health and medicine computerised databases: Medline and Health Plan. Articles were identified according to there being present one of the following keywords in the title or in the abstract: willingness-to-pay, willingness-to accept and contingent valuation. The medical and health literature search was complemented with a literature review on Econlit, the main computerised economics literature database, using the keywords health and one of the following: contingent valuation, willingness-to-pay and willingness to accept. Retrieved titles and abstracts from medical and economics databases were then reviewed to select the articles that reported results from an empirical contingent valuation. Therefore, methodological and review articles, including letters and editorials were excluded. Finally, the bibliography of each collected article was systematically reviewed to identify additional studies of interest.

For the period 1997-2005 a similar, but somehow simplified, approach was deemed necessary because of the much larger number of studies in the area of cost-benefit analysis. Using the same strategy adopted for the period 1984-1996 we obtained 21,262 articles, too many to be processed and selected. Therefore, for the latter period we only retrieved abstracts that contained the expression "contingent valuation" in the title or in the abstract and then we reviewed each abstract to see whether the article included the inclusion criteria, namely to report a CV empirical study. Abstracts of studies were retrieved from both Medline and Econlit. For the Econlit database contingent valuation was associated to the world health in any possible field.

Some of the identified articles report empirical studies on WTP or WTA for mortality risk reduction or health improvement that are not related to health care services. For example, Jones-Lee et al. (1985) investigated the value of reductions in the risk of traffic death surveying a random sample of 1,150 individuals and Shechter and Kim (1991) surveyed individuals living in the Haifa metropolitan area to measure their willingness-to-pay for pollution abatement. These and other retrieved studies used contingent valuation to measure willingness-to-pay (or willingness-to-accept) for decisions affecting people's health that are undertaken outside the health care sector. Although these studies may provide useful experience and evidence to construct a contingent valuation survey for IVF, we decided to focus the literature review on published studies that involve the use of health services and, consequently, to exclude those that are unrelated to decisions made in the health care system.

According to these criteria, we identified 141 studies. For four studies it was not possible to find a copy of the entire article. As a consequence we analysed 136 studies, 32 published between January 1984 and December 1996 and 105 published between January 1997 and December 2005 (table 3.2). The number of contingent valuation studies is growing. The number of published studies in the last two year (2004-05) is higher than those published in the previous two-year period and the growth appears linear or even exponential. Various systematic literature reviews found a similar pattern (Diener et al. 1998; Klose, 1999; Olsen and Smith, 2001; Smith, 2003).

Table 3.2. Contingent valuation studies by purpose and type of problem investigated (1984-2005).

	First Author	Journal	Year of Pub.	Purpose of the study	Problem investigated	Existing/ Hypothetical treatment
1	Thompson	Med Dec Mak	1984	Method/ Demand	Rheumatoid Arthritis	Hypothetical
2	Berwick	Med Care	1985	Method	Diagnostics (antenatal care)	Existing
3	Fisher	J Paediat	1985	Method/ Demand	Paediatric confidential care	Exiting
4	Thompson	AJ Public Health	1986	Demand	Rheumatoid Arthritis	Hypothetical
5	Grimes	BMJ	1988	CBA	Diagnostic (screening for cancer)	Existing
6	Appel	Med Care	1990	Method/ Demand	Diagnostic (contrasting agents)	Existing
7	Donaldson	J Health Econ	1990	CBA	Hospital care vs. nursing home	Existing
8	Reis	J Nursing Admin	1990	Demand	Insurance plans	Not applicable
9	Eastaugh	I J Tech Ass H C	1991	Demand	Blood transfusion	Existing
10	Johannesson	J Health Econ	1991	Method	Hypertension	Existing
11	Pennie	Can J Public Health	1991	Demand	Vaccine for Hepatitis B	Existing
12	Johannesson	J Hypert	1992	CBA	Hypertension	Existing
13	Golan	Med Dec Mak	1993	Method/Demand	Supplementary health insurance	Not applicable
14	Johannesson	J Health Econ	1993	Method	Hypertension	Existing
15	Johannesson	Applied Econ	1993	Method	Lipid lowering	Existing
16	Lindholm	I J Tech Ass H C	1994	CBA	Prevention of cardiovascular diseases	Existing
17	Mills	J Trop Med Hyg	1994	CBA	Bed-net impregnation for malaria	Existing
18	Neumann	Med Care	1994	Method	In-Vitro-Fertilisation	Existing
19	O'Brien	Med Dec Mak	1994	Method	Chronic lung diseases	Hypothetical
20	Donaldson	Health Econ	1995	Method/Demand	Diagnostic (screening fro cystic fibrosis)	Existing
21	Eckerlund	Health Policy	1995	Demand	Health care budget	Not applicable
22	Granberg	Acta Gyn Ob Sc	1995	CBA	In-Vitro-Fertilisation	Existing
23	Miedzybrodzka	J med Genet	1995	Demand	Diagnostic (screening fro cystic fibrosis)	Existing
24	O'Brien	Pharmacoecon	1995	CBA	Antidepressant drug	Existing
25	Osmond	J Paed	1995	CBA	Paediatric sutures	Existing
26	Ross	Med Care	1995	Method	Ambulatory services	Existing
27	Chestnut	Med Dec Mak	1996	Method	Drug for angina	Hypothetical
28	Kartman	Health Econ	1996	Method	Drug for oesophagitis	Hypothetical
29	Kartman	Med Dec Mak	1996	Method	Drug for angina	Hypothetical
30	Ryan	Health Econ	1996	Method	In-Vitro-Fertilisation	Existing
31	Stalhammar	Med Dec Mak	1996	Method	Drug for ulcer	Hypothetical
32	Weaver	Soc Sc & Med	1996	Demand	Quality improvement	Not applicable
33	Asenso-Okyere	Health Policy	1997	Demand	Private Health Insurance	Not applicable
34	Kartman	Health Econ	1997	Method	Drug for oesophagitis	Hypothetical
35	Lee	Health Policy	1997	Demand	Autologous blood donation	Existing
36	O'Conor	J Health Econ	1997	Method	Asthma	Hypothetical
37	Olsen	Health Econ	1997	Method	Cancer, heart interv., helicopter services	Hypothetical
38	Zethreaus	Brit J Ob & Gynaec	1997	CBA	Hormone Replacement Therapy	Existing
39	Donaldson	Birth	1998	Method	Location of giving birth	Existing
40	Lee	Medical Care	1998	Method/CBA	Autologous blood donation	Existing
41	Mathiyazhagan	I J Tech Ass H C	1998	Demand	Community Health Insurance	Not applicable
42	O'Brien	Medical Care	1998	Method	Cancer treatment	Existing
43	Onwujeckwe	Trop Med and Int H	1998	Demand	Ivermectin distribution	Existing
44	Ortega	Cancer	1998	CBA	Prophylactic Eiopoietin	Existing

					alpha treatment	
45	Tambour	Med Dec Mak	1998	Method/CBA	Hormone Replacement Therapy	Existing
46	Chiu	Health Policy	1999	Demand	In-home respite care	Hypothetical
47	Donaldson	Soc Sc & Med	1999	Method	Location of giving birth	Existing
48	Dranitsaris	I J Tech Ass H C	1999	CBA	Adjunct therapy for multiple myeloma	Existing
49	Matthews	Public Health Dent	1999	Method	Periodontal therapy	Existing
50	Onwujekwe	Public Health	1999	Demand	Equity in ivermectin distribution	Existing
51	Sorum	Med Dec Mak	1999	Method/Demand	Prevention of acute otitis	Hypothetical
52	Torrance	Pharmacoecon	1999	CBA	Treatment of chronic bronchitis	Existing
53	Cho-Min-Naing	S Asian J Trop Med	2000	Method/Demand	ICT Malaria Pf/Pv test kit	Existing
54	Clarke	Applied Econ	2000	Method	Mammographic screening	Existing
55	Dranitsaris	Pharmacoecon	2000	CBA	Insuline treatment	Existing
56	Estaugh	I J Tech Ass H C	2000	Method/Demand	Treatment of Von Willebrand's Disease	Existing
57	Liu	Health Econ	2000	Demand	Cold prevention	Hypothetical
58	Narbro	I J Tech Ass H C	2000	Demand	Obesity treatments	Hypothetical
69	Onwujekwe	Trop Med & Int Health	2000	Method/Demand	Re-treatment of Insecticide-treated nets	Existing
60	Papatheofanis	Q J Nucl Med	2000	Demand	PET in suspected lung cancer	Existing
61	Slothuus	I J Tech Ass H C	2000	Method	Arthritis symptom alleviation	Existing
62	Suh	J Am Pharma Assoc	2000	Demand	Pharmacy services	Existing
63	Thomas	J H Res & Policy	2000	Method	Bone mineral density screening	Existing
64	Wagner	Health Policy	2000	Method	Mammographic screen.	Existing
65	Zarkin	J Sub Abuse Treat	2000	Demand	Substance Abuse treatment	Hypothetical
66	Blumenschein	J Health Econ	2001	Method	Asthma management programme	Existing
67	Dalmau-Matarrodona	Health Econ	2001	Method	Home care	Existing
68	Gyldmark	Soc Sc & Med	2001	Method/Demand	Insurance for 4 interventions	Existing
69	Morris	Med Dec Mak	2001	Method	Pneumonia Vaccine	Hypothetical
70	Onwujekwe	Tro Med & Int Health	2001	Method/Demand	Insecticide-treated nets	Existing
71	Onwujekwe	Health Econ	2001	Method/Demand	Insecticide-treated nets	Existing
72	Wagner	Health Policy	2001	Method/Demand	Mammographic screen.	Existing
73	Arana	Health Econ	2002	Method	Influenza vaccine	Existing
74	Bhatia	Health Pol Planning	2002	Demand/method	Insecticide-treated nets	Existing
75	Clarke	Health Econ	2002	Method	Mammographic screening	Existing
76	Forsythe	Health Pol and Plan	2002	Demand	HIV counselling and testing	Existing
77	Nocera	Int of H C Fin and Econ	2002	Method/CBA	Alzheimer Disease	Hypothetical
78	Onwujekwe	Soc Sc & Med	2002	Method	Insecticide-treated nets	Existing
89	Stewart	J Health Econ	2002	Method		Hypothetical
80	Taylor	Health expect	2002	Method	Delivery	Existing
81	Wagner	Health Policy	2002	Method/Demand	Mammography	Existing
82	Whittington	Vaccine	2002	Demand	HIV/AIDS vaccine	Hypothetical
83	Zillich	Pharmac	2002	Method	Asthma	Hypothetical
84	Bhatia	App Ec Lett	2003	Method/Demand	Insecticide-treated nets	Existing
85	Borisova	Health Econ	2003	Methods	Travel time for methadone treatment	Not applicable
86	Dong	Soc Sc & Med	2003	Demand	Community insurance	Not applicable

87	Dong	Health Econ	2003	Demand	Community insurance	Not applicable
88	Foreit	Health Policy	2003	Method/Demand	Reproductive/family planning	Existing
89	Hammerschmidt	I J H Care Fin Econ	2003	Method	Diabetes	Hypothetical
90	Luchini	Health Econ	2003	Method	Cancer, heart interv., helicopter services	Hypothetical
91	Onwujekwue	Health policy plan	2003	Method/ Demand	Insecticide-treated nets	Existing
92	Shiell	Health Econ	2003	Method	Vaccine	Hypothetical
93	Tarasiuk	Sleep	2003	CBA	Obstructive sleeping apnea syndrome	Existing
94	Whynes	J Health Econ	2003	Method	Colorectal screening	Existing
95	Amin	H Res Pol & Syst	2004	Demand	Childhood diarrhoea	Existing
96	Asgary	Eur J Health Ec	2004	Demand	Health Insurance	Not applicable
97	Bissai	Bull of WHO	2004	Demand	AIDS vaccine	Hypothetical
98	Bradford	J Telemed	2004	Demand	Telemedicine	Existing
99	Carlsson	Haemophilia	2004	CBA	Treatment of haemophilia	Existing
100	Dong	H Pol & Planning	2004	Demand	Community-based health insurance	Hypothetical
101	Dranitsaris	Pharmacoecon	2004	CBA	Pharmac. treatment of advances cancer	Existing
102	Greenberg	Health Policy	2004	Method	Prevention of Restenosis	Hypothetical
103	Hammerschmidt	Health Econ	2004	Method	Diabetes	Hypothetical
104	Mataria	J Health Econ	2004	Demand	Quality of care	Not applicable
105	Olsen	J Econ Psychology	2004	Method	Cancer, heart interv., helicopter services	Existing
106	Olsen	Health Policy	2004	Method	Heart operations, hip replacement, cataract	Hypothetical
107	Onwujekwe	J H Serv Res Pol	2004	Demand	Insecticide-treated nets	Existing
108	Onwujekwe	Health Econ	2004	Demand	Insecticide-treated nets	Existing
109	Onwujekwe	Soc Sc & Med	2004	Method	Insecticide-treated nets	Existing
110	Pavlova	App Econ	2004	Demand	Health services	Not applicable
111	Protiere	Soc Sc & Med	2004	Method	Cancer, heart interv., helicopter services	Existing
112	Rheingans	Filaria J	2004	Demand	Lymphatic filariasis	Existing
113	Ryan	J Health Econ	2004	Method	Cancer, heart interv., helicopter services	Existing
114	Ryan	Health Econ	2004	Method	In Vitro Fertilisation	Existing
115	Whynes	Health Econ	2004	Method	Colorectal Screening	Existing
116	Asfaw	Int J H Care Fin Econ	2005	Demand	Health services	Existing
117	Barner	Res in Soc A Pharma	2005	Demand	Pharmacist's counselling for menopausal symptoms	Existing
118	Bradford	Telemedice and e-Health	2005	Demand	Telemedice	Existing
119	De Ridder	Eur J Health Ec	2005	Method	Attention disorder	Hypothetical
120	Dong	Health Policy	2005	Demand	Community-based health insurance	Not applicable
121	Fautrel	J of Rheumat	2005	Cost of illness	Rheumatoid Arthritis	Hypothetical
122	Finkelstein	Am J of Man Care	2005	Demand	Bariatric surgery	Existing
123	Hackl	Health Econ	2005	Method	Red Cross services	Existing
124	Hamelsky	Cephalgia	2005	Cost of illness	Migraine	Hypothetical
125	Ho	Acc Anal & Prev	2005	Demand	Pain-killing pill	Hypothetical
126	Lee	H and Q of Life Outcome	2005	Demand	Pertussis treatment and vaccination	Hypothetical
127	Liu	Health Econ	2005	Demand	SARS Vaccine	Hypothetical
128	Masiye	Malaria Journal	2005	CBA	Malaria vaccination and treatment	Hypothetical
129	Olsen	J Health Econ	2005	Method	Cancer, heart interv., helicopter services	Existing
130	Onwujekwe	Soc Sc & Med	2005	Method/Demand	Insecticide-treated nets	Existing
131	Onwujekwe	Health Econ	2005	Method	Insecticide-treated nets	Existing
132	Sadri	Pharmacoecon	2005	Demand	Inhaled Insulin	Existing
133	Smith	J Health Econ	2005	Method	Hypothetical pill	Hypothetical

134	Suraratdecha	Health Policy	2005	Demand	AIDS vaccine	Hypothetical
135	Takemura	Health Econ	2005	Method	X-ray for gastric cancer	Existing
136	Whynes	Health Econ	2005	Method	Screening (cervical and colorectal)	Existing

The contingent valuation method has received a lot of attention from health economics journals, as it is clearly demonstrated by the high number of studies published in these journals. However these studies remain a marginal component of economic evaluation literature; cost-effectiveness and cost-utility analysis are largely the preferred method for evaluating medical and other health care decisions.

Contingent valuation studies benefit from a multidisciplinary effort. Although the method is rooted in welfare economics, well designed studies also require survey research skills and advise from medical experts. The audience of contingent valuation studies are heterogeneous as well. From the scientific point of view this includes welfare applied economists, health service researchers as well as policy analysts. But as a method to offer guidance on decisions about resource allocation, contingent valuation research is also expected to be of great interest to policy makers, managers, public health officials and clinicians. Consequently, it is not surprising that contingent valuation studies have been published in various types of journals: from public health to general applied economics, from health service research to clinical medicine.

We have classified the literature according to three broad categories of scientific journals: economics and policy, health service research, and medical and public health (figure 3.1). Surprisingly, the first article published in an economic or health policy journal is dated 1990, seventeen years after the paper by Acton (1973) reporting the first attempt to use the contingent valuation method in health care. However, since 1990 the number of studies published in the economics and policy literature has increased substantially and it is now much higher than the number of studies published in public health and medical journals. Data for 2004-2005 clearly illustrates that the trend observed in the '90s has accentuated in most recent years. CV studies mainly appear in policy and, even more frequently, in health economics journals. In only two years 8 studies were published in *Health Economics* and 4 studies in *Journal of Health Economics,* the two most important scientific journals in the field.

The identification of the journal where contingent valuation studies are published is important for two main reasons. First, it is likely that the skills of referees differ across journal types, with economics and policy journals having the most experienced reviewers in the area of applied welfare economics. To the other side, medical journals are much more likely to have experienced reviewers in the specific context of the study but little knowledge on methodological issues. The second reason to monitor where contingent valuation studies are published concerns their audience. Although still at an experimental stage, these studies are aimed at offering guidance to decision making. Publications in policy, public health and medical journals are thus more suitable to inform decision-making as their audience include health care professionals and policy advisors.

However, publications of contingent valuation studies in medical and public health journals, those that tend to have a larger audience and a greater impact on decision making, should raise some concern. The lack of experienced reviewers, as well as the lack of established guidelines on how to conduct these studies, may result in a weak review process and, consequently, in poor quality publications that may affect decision-making at policy, management and clinical level.

The relatively small number of empirical studies is associated with a limited number of researchers committed to the contingent valuation method. Although 78 authors appear in the 33 studies included in the analysis for the period 1984-1996, only 16 of them have published more than one paper and only three have published more than two papers. It is thus possible to assume that very few researchers were mainly focused on this type of studies and that performing a contingent valuation in the health care field was, at that time, occasional for the vast majority of researchers. In that period, the only exception was represented by Magnus Johannesson, a Swedish economist, who is first author in four studies and co-author in other five. Now the situation appears slightly different. About a dozen scientists published more than one article in the period 2004-2005 (Cairns, Donaldson, Dong, Fox-Rushby, Hanson, Onwujekwe, Olsen, Protiere, Ryan, Shackley, Smith). Some of them are also the authors of critical reviews of the existing literature (Olsen and Smith, 2001; Smith, 2003; Hanley et al., 2003). Some of them have participated to the Eurowill project sponsored by the European Commission to test the use of willingness-to-pay to elicit community preferences in 6 EU member states. While until the end of the '90s the scientific interest for contingent valuation studies in the health field appeared limited to very few researchers, in most recent years a more structured community has emerged and a few scientists have clearly made contingent valuation and related techniques their major area of research. These scientists are mainly based in UK and, to a lesser extent, in Scandinavia and France. Interestingly, US researchers show a limited interest in this area of research. We were not able to identify any contingent valuation study published by Italians.

In the period 1984-1996, the commitment of Magnus Johannesson and the Stockholm School of Economics to the use of contingent valuation is witnessed by the relatively high number of studies performed in Sweden. In the same period other countries where these studies have been carried out were the United States of America, United Kingdom, Canada, Australia and Israel. The location of the studies clearly suggests that surveying people about their willingness-to-pay for health care has been mainly attempted in English speaking countries and Sweden. While the concentration of the studies in these countries may reflect a wider diffusion of economic evaluation of health care programmes in general and an easier access to international journals, it cannot be excluded that cultural characteristics of these countries can have contributed to make monetary evaluation of health benefits more acceptable to surveys' respondents and decision-makers.

English speaking countries maintained their leading role in experimenting contingent valuation. In 2005, 45% of studies were performed in UK, Australia or the USA. However, a major novelty emerges from the analysis of the countries where the surveys were performed. Thirty-nine percent of studies published in the 2004-05 period present data collected in developing countries. It appears that contingent valuation is an attractive method to investigate the demand for health care services and to develop co-payment strategies in these countries. Interestingly, few WTP studies performed in low and medium income countries are part of full cost-benefit analyses of specific health care interventions.

3.3.2. Methodological studies

Over the entire 1984-2005 period 55 studies (40%) were methodological as they tested feasibility, validity and reliability of the contingent valuation method (table 3.1).

Contingent valuation has gained importance and prestige in the scientific community but methodological issues are still the focal point of empirical research.

Feasibility and face validity of the contingent valuation method

The first methodology studies focused on testing the feasibility of asking patients or healthy people WTP questions, the feasibility of different value elicitation methods and the association between WTP amounts and some expected explanatory variables. Thompson et al. (1984) and Thompson (1986) investigated WTP for a hypothetical cure for arthritis. In the former study the rate of plausible response was 27%, while in the latter it was 84%. According to the author (Thompson, 1986), the increased response rate of the second study was attributable to improved questionnaire design, to the performance of the interviewers and to having no subjects older than sixty-six.

Johannesson et al. (1991) compared open-ended and take-or-leave-it contingent valuation questions in a mail questionnaire sent to 481 hypertension patients enrolled in a Swedish primary care centre. Patients were divided in two groups. The first sub-sample received and open-ended question in which they were asked to state their maximum willingness-to-pay for their current treatment. The other sub-sample received a yes/no question in which respondents were asked to accept or reject a specified increase in user fees for their current treatment. Usable responses measured as percent of population were much higher with the take-it-or-leave-it question (56%) than with the open-ended question (27%). This study clearly indicates that respondents found the discrete valuation question easier to answer.

In two later studies conducted in the early '90s (Johannesson et al. 1993a; Johannesson et al. 1993b) other value elicitation methods were tested. The former study surveyed hypertensive patients in a primary care centre using a new type of binary WTP questions that allowed for different degrees of certainty with respect to the responses. The overall response rate was about 64% and the item non-response rate on the willingness-to-pay question was about 5%. In the latter study concerning treatment of high lipid levels, a combination of open-ended and closed-ended contingent valuation questions were asked to patients enrolled in a clinical trial. The response rate to the WTP question was 94%, suggesting the possibility of obtaining very high response rates by using the contingent valuation method in clinical trials. Both these studies provided an indirect test of validity as they assessed whether hypothesised theoretical relationship between WTP and some explanatory variables were supported by data. Accordingly to theoretical expectations, the survey on patients with hypertension found that higher prices reduced the demand for the treatment and that a larger difference in perceived health status between treatment and no treatment increased willingness to pay. However, no association was found between socio-economic variables and willingness to pay. Results of the lipid lowering study provided a stronger validity test: increased difference in perceived health status between treatment and no treatment increased WTP, income elasticity was positive and the correlation between WTP and willingness to give up time to take part to the lipid lowering programme was high.

Regression analyses to verify internal validity has always been a basic feature of contingent valuation studies and generally indicate that WTP questions are not answered randomly by respondents (Neumann and Johannesson, 1994; O'Brien and Viramontes, 1994; Kartman et al., 1996a; Kartman et al., 1996b; Ryan, 1996; Stalhammar 1996; Ryan et al., 1997; Zilich et al., 2002). In more recent contributions more sophisticated statistical techniques have become a regular companion of contingent valuation studies. Important theoretical contributions have focused on the search for appropriate techniques for econometric analysis of WTP data elicited in contingent valuation studies (Donaldson et al., 1998). It is now generally recognised that Ordinary Least Square (OLS) estimates are not appropriate because censoring, truncation or the elicitation format used (Maddala, 1983; Long, 1997). For TIOLI questions probabilistic models are generally appropriate (logit and probit). Similarly, when data contain a large proportion of zeros tobit or spike models are expected to perform better than OLS (Donaldson et al., 1998; Long 1997) or to make assumptions more consistent with the nature of data (Kristrom, 1997). In general, most recent studies appear to take statistical analysis more seriously. In addition to using limited dependent variable models; they present various measures to test homoscedasticity, normality of estimated residuals and functional specification (see for example Dong et al., 2003a). Finally, an increasing number of studies use non parametric approaches that allow to avoid to make assumption about the distribution underlying the data sample (Hanley, 2003).

While early studies proved that respondents living in affluent countries, and often with relatively high levels of education, understand WTP questions and provide reasonable answers, more recent studies have tested the feasibility of contingent valuation in developing countries (Amin, 2004; Asfaw and von Braun 2005; Bhatia and Fox-Rushby 2001 and 2003; Dong et al., 2004 and 2005; Mataria, 2004; Onwujeckwe et al., 2004; Onwujeckwe et al. 2005a and 2005b; Masiye 2005; Suraratdecha 2005; Onwujekwe et al., 2008). The method seems to perform well even in these contexts, especially in personal interviews conducted by well trained interviewers and with carefully designed questions. Unsuprisingly, evidence clearly shows that surveys should be designed to fit cultural contexts (Wagner et al, 2000 and 2001; Onwujekwe et al., 2008).

Convergent and criterion validity

In the early 1990s two studies attempted more conclusive validity tests. In addition to investigate the association between household income and WTP, O'Brien and Viramontes (1994) tried to explore the validity of their study by measuring the correlation between WTP elicited by a bidding game and utility values derived from a standard gamble exercise. A relatively high correlation was found between the two measures, indicating that there was some evidence of convergent validity for WTP with preferences measured by standard gamble. It is worth noting, however, that this validity test does not appear very strong because it just suggests that two different ways to measure expressed preferences in a hypothetical situation provide consistent results. This issue has been partly addressed only recently (see below).

The second validity study performed by Chestnut et al. (1996) compared two approaches for measuring heart patients' WTP for changes in their angina symptoms. First, they asked patients to report actual expenditure and perceived angina episodes avoided. From these elements an averting-behaviour measurement of WTP was derived. Second, contingent valuation questions were asked to measure WTP for a hypothetical medical treatment that could avoid increases in angina episodes. Results of this small study (only 35 respondents provided data for the comparison) show that the average WTP to avoid additional angina episodes derived from the actual expenditure questions was comparable to the WTP directly elicited. According to the authors, these results provide a new test of validity of the contingent valuation approach. Unfortunately, the article is quite obscure in many crucial parts so that a complete analysis of the methodology used is not always possible. Moreover, the sample size is very small and some methodological choices are not clearly justified.

Both of these studies tried to produce a convergent validity test. This is a test to measure whether two or more measurement tools, that are expected to measure similar concepts, provide high correlated values (Carmines and Zeller, 1979). In the late '90s, in a study investigating three different health care programmes, Olsen and Donaldson (1998) found that there was a substantial discrepancy between the explicit ranking and the ranking implied from the partial WTP values. More recently, in a study conducted in 6 different countries on 1,240 subjects Olsen et al. (2005) found similar evidence. "The most consistent result of the study is the inconsistency of WTP and explicit ranking in all six countries" (Olsen et al., 2005, 994). The authors claim that their findings seriously challenge the use of contingent valuation to elicit people's preferences in priority setting contexts.

Several studies have tested convergent validity by comparing different WTP elicitation methods. As already reported, Johannesson et al. 1991 compared the open-ended (OE) method with the dichotomous choice (DC) method and Donaldson et al. (1997a) compared the OE with the payment card (PC) method. Both DC and PC approaches to eliciting WTP for health care interventions appear superior to the OE method. With respect of expected determinants of WTP (e.g. that income is positively associated with willingness to pay) and completion rates both the DC and PC approaches are better than OE. In addition, there is concern about face validity of OE questions as they are too hypothetical, they do not reflect the way people generally behave in real markets and they do not provide any guidance to respondents on how to think about their willingness to pay for the good (Donaldson et al., 1997a; Ryan et al., 2004). However, a study on the screening for colorectal cancer has recently provided a different picture (Whynes et al., 2003). In comparison with the open-ended method, the payment scale generated higher evaluations. In addition, it is argued that open-ended questions are closer to how donations are solicited and decided than dichotomous choice questions (Onwujekwe, 2002; Onwujekwe and Uzochukwu, 2004). While people generally take or leave fixed prices in most markets, they are generally asked to give and then it is up to them to define the amount of the donation.

Ryan et al. (2004) have compared WTP estimates generated from PC and DC and have compared stated WTP with preferences elicited from ordinal ranking and cardinal person-trade-offs; they obtained that the DC WTP approach resulted in welfare estimates substantially higher than those produced by the PC approach. They argue that this may be partly explained by "yea-saying". Consistently with DC studies in the environmental economics literature, the DC method results in higher WTP because people bias their answers towards "yea-saying" when this is thought to be the socially desirable response. Neither the DC nor the PC method demonstrated to be convergent with ranking and person-trade-offs exercises. Nevertheless, the authors attributed this result to the fact that preferences are constructed in response to stimuli rather than revealed. Therefore, it can be expected that different methods result in different preferences. These results are partly backed by Shiell and Gold (2003) who have shown that preferences are not well formed and ready to be revealed in contingent valuation surveys. They also found that the payment card method failed to reveal respondents maximum WTP and provided vague estimates of preferences. In a major study conducted in the USA to compare 7 different value-elicitation methods it was obtain a similar result: 4 out of the seven methods did not produced data consistent with different indirect utility functions and the three methods that generated inconsistent estimates with the other were OE and PC (Cameron et al, 2002). It can be argued that these two methods may lead respondents to think about CV questions in rather different ways (Hanemann, 1996)

Coming back to healthcare studies, another important convergent validity test was performed by comparing CV method and travel cost method for improved access to mammographic screening in Australia (Clarke, 2002). The travel cost method measures revealed preferences and estimates the benefits of the good by using travel costs to have access to it. Based on information collected through telephone interviews of 372 women, the study by Clarke (2002) estimated that WTP based on the CV method was significantly higher than the WTP based on the travel cost method.

Tests of convergence provide valuable evidence on the validity of the CV method. Even if it is accepted that different cognitive processes may results in different preferences, a certain degree of convergence is to be found to make the contingent valuation method credible. If two elicitation approaches provide two unrelated estimates, at least one of them is not valid (that is it does not measure what it is designed to measure). However, a better validity test of the CV method is to compare hypothetical with real WTP.

To our knowledge three studies performed such a criterion validity test in the health care field (Blumenschein et al., 2001; Bhatia and Fox-Rushby, 2003; Onwujekwe and Uzochukwu, 2004; Onwujekwe et al., 2005). In the first study subjects received either a dichotomous choice WTP question or were given the opportunity to actually purchase an asthma management program (Blumenschein et al., 2001). In the hypothetical group, 38% of subjects stated that they would purchase the program at the stated price, but only 12% of subjects in the real group purchased the program. However, if in the hypothetical group only responders who stated to be "definitely sure" about their willingness-to-pay were considered "purchasers", the fraction of hypothetical (14%) and real purchasers (12%) would be very similar. This study clearly indicates that the DC method overestimates WTP. However, it also suggests that it may be possible to correct this overestimation by sorting out the "definitely sure" and "yes" responses.

The second study was carried out in Gujarat (India) and involved three hundred households in 20 villages (Bhatia and Fox-Rushby, 2003). Respondents were first interviewed twice and then asked whether they would be willing to buy a treated mosquito net for a fixed price (equal to the modal value revealed in the interviews). The authors found that at an aggregate level there was no statistically significant difference between hypothetical WTP and actual demand. However, there was considerable variation among individuals. The authors concluded that CV-based WTP estimates are robust at the population level but less so at the individual level.

The third study compared three elicitation CV methods with actual willingness-to-pay for insecticide-treated bed nets in Nigeria and investigated the factors that cause divergences (Onwujekwe et al., 2005b). For all the three question formats there were divergences between stated and actual WTP: 69.4% of respondents in the bidding game group, 78.8% in the binary with follow-up group and 48.8% in the structured haggling group behaved differently from what they stated in the contingent valuation survey. Actual behaviour diverged from stated WTP for both the people willing to pay (who then did not purchase the net) and those unwilling to pay (who then did purchase the net). The authors found that there were genuine causes of divergence between stated and actual WTP across the question formats and argued that it is important to detect whether the divergence in WTP occurs because of bias or because of legitimate changes in consumers' values or circumstances.

The same study investigated an issue that is probably even more crucial to understand the validity of the contingent valuation method in health care: the comparison of stated and actual altruistic WTP (Onwujekwe and Uzochukwu, 2004). First of all, the authors investigated the appropriate format to ask altruistic WTP questions. In general, it can be argued that question formats that better mimic price-taking in a particular context will generate more valid WTP estimates. However, according to the authors it is unclear whether the same question format should be used to value the three aspects of WTP (use, option and altruistic values). The dichotomous choice (yes/no to a proposed value) generally resembles the real market decisions of individuals for valuing use or option values, but it may not resemble the people's normal altruistic decisions for goods and services. In general, dichotomous choice, bidding game and payment card formats do not reflect how individuals make altruistic contributions. These "payments" are usually made freely and voluntarily in an open-ended way.

To determine the best valid estimates of altruistic WTP the authors randomized the respondents either to the take-it-or-leave-it (TIOLI) with follow-up format or to open-ended (OE) questions. They then compared elicited WTP to provide insecticide-treated nets to low income households in Nigeria to actual voluntary contributions. More than 57% of the respondents were hypothetically willing to pay for altruism in both groups. A total of 27% and 33% of the respondents with positive hypothetical WTP in the TIOLI with follow-up and OE actually contributed. This study shows that only about one third of respondents willing to pay then actually paid the stated amount. It is also shows that the open-ended format elicits better valid estimates of altruistic WTP than the dichotomous choice format.

Overall, there is too little evidence available on the validity of the different formats to elicit WTP for health services. So far the following provisional conclusions can be formulated. First, different elicitation methods provide different stimuli and thus produce different results. Consequently, there is no "right" method to elicit WTP.

Second, there is evidence of "yea-saying" and this tends to generate overestimates of WTP when the take-it-or-leave-it format is used. The most precious experiments, which are those that compare stated WTP to actual WTP, clearly suggest that the former is larger than the latter. This is very important from a policy perspective and suggests that it may be useful to measure the degree of certainty of the respondents. Finally, to provide valid measures, altruistic WTP may require special elicitation formats that reflect the way in which altruistic payments occur in reality.

Reliability

In assessing reliability, it is necessary to investigate whether measuring the same object or the same phenomenon with the same measurement instrument will give similar results (Carmines and Zeller, 1979). Correlations between duplicated or reproduced measurements of the same object or phenomenon, using the same instrument, need to be calculated. This duplication can be carried out either by the same observer at different times (test-retest reliability), or by different observers simultaneously (inter-rater reliability).

The evidence on the reliability of willingness-to-pay elicited using the contingent valuation method is growing. The first study included in our review, that used open-ended questions, performed a test-retest reliability check on about 50 patients and found a correlation of 0.25 (Thompson et al., 1984). Studies performed in the '90s (mainly bidding games) suggested moderate-to-high reliability; reliability test-retest coefficients ranged between 0.5 and 0.8 (O'Brien and Viramontes, 1994; Cho et al., 2000; Sorum, 1999). More recently, Onwujecke et al. (2005) retested bidding game, closed-ended and structure haggling questions one month after the first face-to-face interview and got relatively low Pearson's coefficient of correlation (0.3-0.5). In this study structured haggling performed satisfactorily but it was not clear which format was superior. In a larger study in Africa conducted with face-to-face interviews, Dong et al. (2003) found that the reliability of both the dichotomous choice and the bidding game formats were moderate to good (Pearson and Spearman correlations between 5.9 and 7.0). In addition, the study shows that the bidding game format was more reliable than the closed-ended format.

To our knowledge the best evidence on reliability comes from the Nigerian study on insecticide-treated nets carried out at the London School of Hygiene and Tropical Medicine (Onwujekwe et al., 2004). Inter-rater reliability was evaluated by having two sets of interviewers administer questionnaires to randomly selected household heads. Inter-rater reliability coefficients were high at 0.77, 0.75 and 0.74 for the bidding game, the binary with follow-up and the structure haggling technique, respectively. Test-retest reliability was investigated by repeating interviews one month after the initial survey. Test-retest coefficients were low-to-moderate, ranging from 0.4 to 0.56. According to the authors, however, the lower coefficients in the test-retest reliability were due to the influence of factors affecting demand that had changed in the intervening period. This study also shows that the three different question formats have similar levels of reliability.

Random assignment of bids, range bias and prominent numbers

Four studies published recently investigated specific technical issues related to the choice of numbers to present in contingent valuation questions. Takemura et al. (2005) investigated the influence of poor random assignment of bids in the discrete-choice WTP elicitation format. They show that if the survey fails to assign the bid randomly according to characteristics that are expected to be correlated with WTP, results are biased. It is thus necessary to check if bids are randomly assigned to the respondent's characteristics and, if they are not, to adjust their interaction effects.

The other two problems are specific to the payment card format, that is the method where respondents are presented a card with possible WTP values and then they are asked to state if they *are* or they *are not* willing to pay for each of them (Whynes et al., 2004; Whynes et al., 2005). Probably, the most disturbing problem concerns the existence of range bias (Whynes et al., 2004). The authors report the findings of a contingent valuation survey of colorectal cancer screening, wherein different subjects were provided with payment scales of different lengths. The long scale presented values up to UK £ 1,000, while the short scale version extended only to UK £ 100. It emerged that the long scale produced a mean WTP more than 30% higher than that resulting from the short scale. The authors concluded that their findings are strongly supportive of the existence of range bias in payment card instruments – scales which extend to higher values and generate even higher values.

The other problem was discovered by observing the distribution of WTP values in three contingent valuation studies that elicited WTP from samples drawn from the general population of east-central England (Whynes et al., 2004). The studies elicited WTP using a payment scale, completed without supervision (self-administered). Subjects were asked to encircle their maximum value out of 29 different values, having obtained these by placing ticks against amounts they were sure they would pay and crosses against amounts they were sure they would not. The majority of the 3,000 respondents indicated one of a limited number of values from the range available to them (£ 1, 2, 5, 10, 20, 50, 100). These "prominent numbers" has been observed previously in circumstances where respondents were uncertain about their choices and are suggestive of bias in contingent valuation.

Scope sensitivity, "warm-glow" and free riding

In the '90s another major problem, widely debated in environmental economics, was addressed in the health care contingent valuation literature: scope sensitivity. The scope effect refers to the fact that, to be theoretically and intuitively valid, the contingent valuation method has to be sensitive to changes in scope of the programme being evaluated, i.e. that WTP increases with the size of benefits. An often mentioned study illustrates the problem. Boyle et al (1994) found that the average WTP to prevent 2,000 migratory birds from dying in oil-filled ponds was as great as that for 20,000 or 200,000 birds from dying. Marginal WTP is expected to decrease as the units of benefit of the programme increases. However, a drop to zero in WTP for additional benefits is hard to explain as the expression of a consistent, rational set of choices (Arrow et al., 1993).

Kartman et al. (1996b) tested the scope effect using a split sample approach. Different sub-samples were asked to answer a yes/no WTP question for different probabilities of successful treatments for reflux oesophagitis. Therefore, in this context (as in most of health care programmes) the size of the commodity being evaluated was measured in terms of probability of success. In this study the contingent valuation method proved sensitive to changes in scope in that the higher the probability of success, the higher the elicited WTP. Similarly, Neumann and

Johannesson (1994) consistently found that WTP for IVF increased with the probability a success (measured in terms of at least one delivered baby).

A few recent studies have further investigated whether the contingent valuation method is sensitive to change in the scale of the programme. Liu et al. (2005) found that the WTP of two samples of the Taiwanese population were sensitive to the scope of the risk reduction of SARS. WTP was positively correlated (i) to the amount of risk avoided by the intervention, (ii) to the mortality rate attributable to the disease, (iii) to the duration of the efficacy of the preventive measure. Differently, a contingent valuation survey conducted in Uganda to understand the determinants of personal demand for an AIDS vaccine showed that respondents were little sensitive to the efficacy of the vaccine (Bishai, 2004). In a sample of more than 1,000 individuals the fraction of "yes" responses was not different for a 50% or a 95% efficacy vaccine. Respondents were not scope sensitive.

Greenberg et al. (2004) found slightly different evidence in a study conducted in the USA among patients enrolled in two percutaneous coronary interventions trials. The study measured patients' WTP to reduce the risk of restenosis and repeat revascularisation. The baseline scenario described a 30% probability of repeat revascularisation. The median WTP was US $1162, US $ 366 and US $273 for 30%, 20% and 10% absolute risk reductions, respectively. Here respondents were scope sensitive as they revealed higher WTP for higher risk reductions. However, they showed that they were willing to pay an additional amount of almost US $800 to reduce the risk from 10% to 0% and only US $ 87 to reduce the risk from 20% to 10%. This study suggests that patients are willing to pay higher amounts for therapies that completely eliminate the risk of a negative event. One possible explanation may be that many individuals have a poor appreciation for numerical differences in magnitude (Greenberg et al., 2004).

Overall these studies suggest that respondents may have difficulties with fully appreciating numbers, especially if they refer to probabilities (Hammit & Graham, 1999). Our personal experience with people, even with highly educated ones, is that they rarely understand probabilistic reasoning. Obviously, this is a major problem with contingent valuation; monetary valuation should reflect the magnitude of the benefits to be usable and meaningful. However, it is worth noting that these difficulties with numbers and probabilities occur also when people make real choices. The fact that people have serious problems with probabilistic reasoning poses a serious challenge to all economic analysis based on rational choices, not only to contingent valuation studies.

But scope insensitiveness may have a different cause. It is argued that stated willingness-to-pay may be insensitive to variations in scope because respondents do not report real economic preferences, but rather derive moral satisfaction (warm glow) from the act of giving per se (Kahneman & Knetsch, 1992; Hackl & Pruckner, 2005). Therefore, when asked WTP questions, respondents would signal their "warm glow" disregarding the scope (size) of the intervention. We are aware of one study only that intentionally tested the warm glow effect for health care services. Hackl and Pruckner (2005), through a contingent valuation survey of Red Cross services in Austria, found no evidence of the warm glow phenomena. Additionally, they did not find any evidence of respondents' strategic behaviour (that is to give false signals). Both results were consistent across different payment vehicles (donations and insurance premiums). New studies are required to test the hypothesis of warm glow effects in health care; nevertheless, the existing evidence appears in favour of the validity of the method as far this effect is concerned.

Three recent studies have explored the cause of scale insensitivity of WTP and have shown that it tends to decrease as the size of the health benefit being valued increases (Smith, 2001; Yeung et al., 2003; Smith, 2005). They suggest that this may be due to an increasing relevance of the budget constraint as the value of the programme increases. As the benefit generated by the programme increases, WTP for that benefit rises and consequently the budget constraint becomes more determinant. Accordingly to standard economic theory, more of a good is consumed the more utility is gained, but each successive increment in the good yields lower amounts of additional utility (because of diminishing marginal utility). Consequently, in the indirect estimation of utility WTP should also increase at a decreasing rate for successive increases in health status (Smith, 2005).

Starting and order bias

The Nigerian study on insecticide-treated nets provides uncertain evidence on another critical feature of contingent valuation: the possibility that, when asked more than one WTP question, respondents tend to be influenced by the initial proposed value (Onwujekwe and Nwagbo, 2002). This problem is a cause of systematic errors in the elicited WTP and is normally called starting-point bias. O'Brien and Viramontes (1994) found no association between starting bids and final bids when they analysed mean WTP adjusted by income and health status. However, in a different analysis that considered median values there was some evidence of a starting-point bias, but this did not reach conventional statistical significance (p=0.07).

Stalhammar (1996) has provided evidence on this problem by conducting a bidding game with 82 patients affected by duodenal ulcer or reflux oesophagitis. He found that the average WTP among the 42 patients who started at the low bid and the 40 patients who were initially offered a high bid were 70 Swedish Crowns (SEK) (C 60-79) and 289 SEK (CI 214-364), respectively. Consequently, the study indicated the presence of a relevant starting bias. This bias was found in other studies (Chestnut et al. 1996; Liu et al., 2000; Dong et al., 2003a). The study by Dong et al. (2003a) performed in Burkina Faso provides strong evidence in favour of the existence of a starting point bias in the bidding game process. The starting price offered to respondents was significantly associated to the elicited WTP. The higher the initial bid, the higher the elicited WTP. In a more recent study for telemedicine services Bradford et al. (2004) did not find similar evidence. However, the study may have had insufficient power to detect the effect due to the limited sample size (126 answers).

The study by Kartman et al. (1996b) on reflux oesophagitis also investigated the presence of question order effect: changing the sequence of questions in a multiple valuation questionnaire results in significant differences in elicited WTP. To test this potential shortcoming of contingent valuation, patients were randomly assigned to one of six possible sequences of questions. Whether the probability of accepting a certain proposed WTP was affected by the question order was then assessed in logistic regression analyses by using dummy variables. The results of the regression analyses did not prove any impact of question order on WTP.

A few studies have further investigated whether the order in which the programmes (goods) are presented to respondents has an influence on the WTP values obtained. In a study presenting three health care programmes, Stewart et al. observed ordering effects in the ranking of the programmes; in the proportion of zero values reported and in the WTP for one of the programmes (Stewart et al., 2002). According to the authors, the best explanation for the ordering effects is one of warm glow, whereby the first programme in any sequence captures much of the utility associated with giving. Ordering effects are confirmed by De Ridder and De Graeve (2005) and, only to a limited extent, by Carlsson et al. (2004). In the former study a split sample was used to test order bias. The results show that WTP for a new drug varies according to the order in which the alternatives are presented. Respondents stated a higher WTP for the new drug only if the standard therapy (the comparator) was presented first. The latter study, based on a representative sample of the Swedish population, found that there was an order bias for one of the two treatments for which WTP was elicited.

Scenario description

Many healthcare technologies have several effects on the individuals' health. In addition, these effects are not uncertain and thus require probability reasoning to be appreciated. For cost-benefit analysis they should theoretically be valued in a single scenario covering all relevant effects; this approach is called holistic measurement. However, a few studies presented decomposed scenarios in which respondents stated their WTP for each individual effect separately (O'Brien et al., 1995; Dranitsaris, 1997; Luchini et al., 2003). This approach is often preferred because it poses a smaller cognitive burden to respondents. Holistic WTP measures can be too complex especially when a large number of health effects have to be valued at once. As the capacity to process information appears to be limited, a holistic valuation of a technology with several effects can result in being too cognitively demanding. Often, respondents focus on a few effects (most likely to be those presented first).

An early contingent valuation study that measured WTP for an ultrasound in normal pregnancy provided evidence for the existence of sub-additivity: the sum of WTP for each individual effect measured separately exceeded WTP elicited for the holistic scenario (Berwick and Weinstein, 1985). Sub-additivity was confirmed by a recent study by Hammerschmidt et al. (2004) who tested the two approaches on diabetic patients. The authors also suggested a theory-based aggregation of individual effects that could considerably reduce the degree of overestimation generated by the aggregation WTP elicited in decomposed scenarios.

The complicated issue of how to describe the scenarios for WTP questions have been investigated by a group of French health economists who took part of the Eurowill project. This project was designed to assess the feasibility of the CV method as a tool for measuring preferences of members of the general public about health care programmes (Protiere et al., 2004). The project investigated three health programmes: more heart operations, a new breast cancer treatment and helicopter ambulance services. The authors show that performing separate evaluations of different programmes result in WTP estimates that are different from those derived from joint evaluations (that is when more programmes are assessed in the same study) (Luchini et al., 2003). They conclude that separate estimations may lead to misspecifications, as the estimations cannot take into account the fact that joint evaluation exogenously provides a reference structure to the respondent which affects the estimates of WTP for each programme. In another study using the same dataset the authors investigated the effects that providing additional information to respondents had on their WTP values. Although the statistically significant level of 5% was not reached (probably due to the limited power of the study), the value associated with some additional "neutral" information on the process of care was positive. In addition, when the scenario was complemented by unambiguously "positive" information, the increase in the mean WTP became statistically significant. Interestingly, WTP for all three programmes tended to be significantly higher for respondents who were provided additional positive information about only one of the programmes. The fact that providing additional positive information on one programme affected WTP values for the other two programmes needs to be explained. It may be that respondents who received positive signals about a health programme tended to interpret such information as a signal about a general improvement in the quality of health care in general. Alternatively, the cause of this unexpected result may be attributed to a type of "starting-point" bias whereby the elicited WTP values for the two programmes for which additional information was not given simply derived by the previous evaluation. The first programme presented, whatever the amount of information it contained, was a reference point for the other two. Therefore, if the first programme contained more information and thus had a higher WTP, the subsequent programmes would be better evaluated simply because their assessment followed one that was better perceived.

These studies indicate that any attempt to measure preferences for health programmes in the general population should present respondents with descriptions of care containing attributes which are "legitimate" (from the decision maker's point of view) and, potentially utility-generating. Overall, the description of the scenarios to be presented in contingent valuation surveys is probably the most critical element. So far, it has received limited attention and often empirical studies do not carefully design this element of the questionnaire (Diener et al., 1998; Olsen and Smith, 2001). Future studies should further investigate this issue and should identify devices to increase the validity of contingent valuation in this respect.

3.3.3. Non cost-benefit studies providing inputs to decision makers

In the period 1984-1996, seventeen articles (56%) openly used the contingent valuation method to offer guidance to policy, managerial or clinical decisions (table 3.2). However, not all these articles make reference to costs of the programme being evaluated and most of them used the methods outside the framework of cost-benefit analysis. Miedzybroclzka et al. (1995) elicited WTP for two methods of carrier screening for cystic fibrosis from women attending a Maternity Hospital Clinic in Aberdeen. They found that the stepwise method (the mother is tested first and if she is positive the test is administered to the partner) was preferred by more women than the couple method (both future parents are tested simultaneously), although the intensity of the preference (the mean WTP in the two groups) was similar. On the basis of this evidence and on the result of a randomised trial showing that the stepwise method of screening was associated with less anxiety and false reassurance, the authors concluded that step-wise screening was the better approach for implementation. In this study WTP was an instrument used to measure the intensity of preference and was not used in a cost benefit analysis.

Two older studies used the contingent valuation method to investigate WTP for changes in the health care system. Golan and Shechter (1993) elicited willingness-to-pay for a programme that would allow interested patients to receive a variety of services otherwise unavailable through the statutory Israeli health care system. Under the programme; patients would be able to choose a consulting physician or surgeon from among the hospital's senior staff, and to schedule operations or at least reduce waiting times. The availability of these options would be allowed by an extension of senior doctors' working time. It was found that the WTP elicited from a random sample of Haifa residents was reasonable and that as a policy implication a complementary health care programme was favourable. However, no cost data of the programme was reported in the article; it was only mentioned that a parallel study on the supply of senior physicians suggested that adequate compensation could be generated by the activation of the programme.

A similar study was carried out by Eckerlund et al. (1995) to analyse how much Swedish residents are willing to spend in the form of taxes for health care. WTP was elicited by the means of a take-it-or-leave-it contingent valuation question that was included in a telephone survey of a national random sample of 1260 households. Respondents were asked if their households were prepared to pay X more per month in order to improve the medical care in their own county. Five possible answers were given: "yes, definitely"; "yes, probably"; "no, probably not"; "no, definitely not"; "don't know". Using a conservative interpretation of the yes answer to the proposed bids (only respondents who answered "yes, definitely") mean WTP was estimated at 284 Swedish Crowns (SEK) per month (€1 = SEK8), while in the standard interpretation ("yes, definitely" and "yes probably") mean WTP was estimated at 635 SEK per month (health care expenditure per household in Sweden was about 2600 SEK per month at that time). This study interpreted results as a referendum. The authors report that the majority of respondents were prepared to increase the tax payment to health care, although the size of the increase appeared relatively modest (50% of the sample voted in favour of an increase in the tax payment of SEK 64 and SEK 237 per month using conservative and standard yes answers, respectively).

In the period 1997-2005 63 studies (43%) provided management or policy evidence and were not purely methodological. Most of them are not cost-benefit analysis as they investigated the determinants of the demand for specific services or for quality of care characteristics. It is interesting to notice that many studies conducted in developing countries were focussed on willingness-to-pay for personal or family's use of services and aimed at understanding if co-payment schemes were feasible and affordable by the population (Amin and Khondoker, 2004; Bishai et al., 2004; Asfaw and von Braun, 2005; Asgary et al., 2004; Suraratdecha et, 2005; Onwujekwe, 2004;

Onwujekwe, 2004, 2005a, 2005b). It appears that in developing countries, and also to a certain extent also in affluent economies, contingent valuation is mainly used to investigate the characteristics of demand for health services, to set prices and, more generally, to improve market knowledge (see for example the large study on community health insurance in Burkina Faso; Dong et al. 2003a, 2003b, 2004 and 2005).

The paper by Weaver and colleagues on WTP for child survival in Central African Republic is somehow representative of this type of studies (Weaver et al., 1996). The survey addressed two major issues: a) to what extent the user fee finances the costs of quality improvement and b) whether a uniform program could be implemented throughout the country or should the user fees be different across health regions. For each element of the hypothetical quality improvement programme the authors could estimate the percentage of the population that was willing to pay the costs and could assess that WTP in rural and urban areas was substantially different. According to these results the authors concluded that the proposed national user fee/quality improvement program was feasible and appropriate.

More recently, similar studies were conducted in Iran to assess if health care insurance currently operating in urban areas can be expanded to rural areas (Asgary et al, 2004), in Burkina Faso and Ethiopia to measure WTP for community based health insurance schemes (Dong et al. 2004 and 2005; Asfaw and von Braun, 2005), in Uganda to assess the factors affecting demand for an HIV/AIDS vaccine among adults (Bishai et al, 2004), in Bulgaria for measuring population attitudes towards paying for quality improvements of and quick access to public health care facilities (Pavlova et al., 2004) and in Nigeria to elicit WTP for insecticide-treated bednets (Onwujekwe, 2004; Onwujekwe 2005). Other studies that mainly investigated determinants of demand for health care were conducted in India, Thailand, Haiti, Zambia and Palestine. Indeed, the contingent valuation method appears increasingly used to investigate the characteristics of personal demand for services and to estimate the impact of user fees or private health insurance, rather than to assess whether government sponsored programmes produce net economic benefits.

3.3.4. Cost-Benefit studies

Between 1984 and 1996 seven published studies (21%) can be classified as cost-benefit. However, most of these studies do not provide accurate cost analyses of the programme being evaluated. The Donaldson (1990) paper was the first contingent valuation study published in a health economics journal. This British study attempted to determine WTP for two alternative publicly-provided goods, namely continuing-care for elderly people in either hospital or National Health Service (NHS) nursing homes. Respondents were relatives of residents in both types of accommodations. After having provided them with a factual summary comparing hospital and NHS nursing home accommodation, relatives were asked which of the two types of care they preferred. They were then told how much per week the government spends on a continuing-care hospital bed and a NHS nursing-home place (British £215 and £225 respectively at 1985 prices). They were then asked whether they thought a place in the preferred type of care was worth more than the cost of the other. Those who replied "yes" were asked how much more their preferred place was worth and those who gave a negative answer were asked how much less it was worth. Finally, for any valuation over £215 per week (the cost of the cheapest accommodation), respondents were asked whether they would be willing to accept the tax increases relative to such valuation. The household increase in taxes was calculated from a "ready reckoner" made available to interviewers. Seventy-one per cent of respondents provided valuations which could contribute to the analysis. Results show that the group that preferred NHS nursing-home care could potentially compensate the group which preferred hospital care and still remain better off. The technique used in the study is very interesting as it is an attempt to measure WTA and WTP to quantify benefits and costs, respectively, of a change in resource allocation (Gafni, 1991).

To elicit WTP and WTA Donaldson (1990) used a sample of relatives of residents in NHS hospitals and NHS nursing homes and not the residents themselves. Similarly, Mills et al. (1994) surveyed two local informants (headman or religious leader) in 53 villages in Gambia to ask about WTP for insecticide for bednet impregnation and preferred means of paying. During informal interviews the Medical Research Council fieldworkers administered a questionnaire that included an open-ended question on what respondents thought would be the maximum and the minimum amounts compounds would be willing to pay for the insecticide. The comparison of stated WTP with the likely cost of impregnating nets indicated a substantial gap between the two, the actual cost being far greater in the vast majority of the villages. However, the study was probably affected by a strategic bias, as it is likely that respondents understood WTP in the hope that subsidies could be forthcoming.

O'Brien et al (1995) used the contingent valuation method to assess the value of a new antidepressant as an alternative to four products belonging to the same pharmacological category, with different prices, but the same adverse effect profile. From a published meta-analysis of controlled clinical trials, they identified seven adverse effects and asked patients to express their WTP for a new drug that reduced each adverse effect by the specified probability. However, the authors could not derive a precise estimate from the WTP questions because there was uncertainty about the aggregation of WTP for multiple risks. Consequently, they reported that WTP was expected to be between Can $ 36 per month (the highest mean WTP among the 7 elicited values) and Can $ 118 (the sum of the 7 mean WTP values). They then compared this range with the additional costs required to substitute the old treatments with the new one and concluded that the net benefit of the new antidepressant (WTP minus cost) was greater when it was compared with the two most inexpensive treatments, but it was found ambiguous.

Osmond et al. (1995) performed an economic comparison of three methods commonly used to repair paediatric facial lacerations: non-dissolving sutures, dissolving sutures and tissue adhesive. First, they performed a cost analysis taking into account equipment utilisation, health care workers' time and loss of both parents' income for follow-up visits. The authors reported that the cost analysis was based on a healthy child being treated for a simple facial laceration in a tertiary-care paediatric hospital emergency department. According to the cost study, their conclusion was that the tissue adhesive was cheaper than both dissolving and non-dissolving sutures. Second, they approached a convenient sample of thirty parents who were visiting the emergency department with a child who had a problem other than a laceration. The sample was provided with an outline of the three repair techniques and with a description, based on previous studies, of the cosmetic outcome, complications, and level of pain to be expected with each. It was then asked parents to name the first, second and third choice of treatment method they would prefer for their child and how much they would be willing to pay for their first and second choice, if only their third choice was provided by the health care system. Twenty-seven parents named the tissue adhesive method as their first choice. Median WTP was Can $40 and Can $25 for tissue adhesive and dissolving sutures, respectively. Both cost analysis and WTP show that the tissue adhesive method was superior to the other two, so that the authors could conclude that this method was the preferred one of closure paediatric lacerations. It should be noted, however, that performing a cost analysis on just one case is clearly unsatisfactory, that the assumption that both parents lose income for the follow-up visit (required for non-dissolving sutures) appears questionable and that it is likely that the analysis suffered from double counting as WTP is expected to include parents' valuation of their loss of income because of follow-up visits.

Surprisingly, out of the 102 studies published in the 1997-2005 period only twelve can be considered full cost benefit analysis (11.4%). In the clinical literature, from the last two years of the review we identified only two CBA based on contingent valuation. The paucity of cost benefit studies suggests that, in spite of the increasing popularity of the contingent valuation method, scientific interest for its use in the context of cost-benefit analysis is modest. As suggested earlier, the contingent valuation studies recently published either address methodological issues or aimed at investigated the determinants of demand focussing on the user value of services. However, it is also possible that in recent years more CBA studies than methodological ones have been missed. This is because methodological studies general apper in economics journals and "substantive" studies appear in medical and public health journals; and in medical journal it may be more frequent that the term "contingent valuation" is not used in abstracts and titles.

The first of the two latest published studies considered in this review concerned prevention and treatment of lymphatic filariasis (LF) conducted in Haiti (Rheingans et al., 2004) and reports both WTP estimates and provisional cost data of the programmes being evaluated. The authors assert that the estimated WTP for LF prevention in the surveyed community is likely to exceed actual medical expenses and productivity losses. However, details are not given and it is not appropriate to include productivity costs if benefits are calculated through a contingent valuation eliciting WTP for preventing the disease. As it is likely that respondents take into account productivity losses when they provided their answers, productivity losses are counted twice in the analysis. Although a formal quality assessment of the reviewed studies is not presented, it should be noted that this is one of the most obscure and most methodologically questionable studies of the entire sample.

Fortunately, the other cost-benefit study is clearly presented and appears methodologically sound. Carlsson et al. (2004) selected a representative sample of 1080 Swedish adults, mailed them information about two different treatments for Haemophilia and asked them in a telephone interview to answers to two dichotomous choice (yes/no) WTP questions. Different respondents were randomly offered different bids ranging from 0.71 € to € 130. The mean estimated WTP was € 39 for the on-demand treatment and € 65 for prophylaxis treatment. Results were robust although it was detected a bias due to the order in which the two treatments were presented. We would like to point out that WTP for treating haemophilia was elicited from a sample representative of the general population. The survey asked whether people were willing to pay a specific amount so that patients with severe haemophilia would obtain on-demand treatment and another specific amount for prophylaxis. Therefore, the elicited WTP included the altruistic component.

The authors performed a cost benefit analysis as they compared the estimated average WTP per Swedish person (€ 39 to €65 for on-demand and prophylaxis treatment, respectively) to their share of the total cost of treating all patients with severe haemophilia in Sweden (€ 1.97 and € 5.56 for on-demand and prophylaxis treatment, respectively). On the basis of these estimates, the authors argued a) that the study gives support for both treatments strategies (compared with no treatment), since the estimated mean WTP exceeds costs of treatment b) gives firm support for prophylactic treatment, as the additional cost per taxpayer of prophylactic treatment was covered more than sevenfold by the additional WTP. Although this study was not available at the time the contingent valuation presented in this thesis was undertaken, the two studies have a very similar design as both elicit WTP from the general population, estimate the costs of providing a national programme and compared costs and benefit to assess the net economic benefit of the programme.

3.3.5. Health care issues investigated by contingent valuation studies

Table 3.2 also reports specific aspects investigated in the health care contingent valuation literature. In the 1984-1996 period twenty-five studies could be grouped in 6 broad classes. The largest class comprises studies where respondents are presented with hypothetical treatments. All of them are methodological and refer to chronic diseases.

Five studies focus on existing programmes to prevent cardiovascular diseases and refer to pharmacological treatments to lower lipid levels and hypertension, as well as to a community-based primary care prevention programmes that involved the mobilisation of health workers, schools, voluntary associations, educational institutions, food producers, the retail trade, and the media (Lindholm et al 1994). One of these studies, published in a medical journal, was carried out in the context of a randomised clinical trial and compared dietary and obese drug treatment in obese men with mild hypertension (Johannesson and Fagerberg, 1992). In the 1984-1996 period, all the studies on existing programmes to prevent cardiovascular diseases were carried out in Sweden.

Five studies concern diagnostic procedures and highlight the merits of contingent valuation in capturing benefits that cost-effectiveness analysis can hardly consider. Two studies elicited WTP for different methods of antenatal carrier screening for cystic fibrosis (Donaldson et al 1995; Miedzybrodzlca et al 1995), while the other three investigated the value of a low osmolality contrast media that reduces risk of minor side effects (Appel et al., 1990), the value of ultrasound in normal pregnancy (Berwick and Weinstein, 1985) and the value of a negative cervical smear (Grimes, 1988).

The low osmolality contrast study asked patients to imagine a situation in which they required to have a radiography involving intravascular injection of a contrast agent. Patients were presented specific possible side effects of a low and a high contrast media and were provided with the risk of having each side effect with each contrast media. Of the 95 patients (out of 100) who completed the study questionnaire, a majority were unwilling to pay the extra cost of USA $ 50 to reduce the risk of minor side effects. The authors concluded that the use of the new contrast media was questionable. Gafni (1991) criticised several aspects of this paper, including the fact that it did not present the problem in the context of purchasing an additional insurance policy. He further criticised the choice of a convenient sample of patients (clearly belonging to low social classes) as it hindered any reasonable generalisation. The ultrasound in normal pregnancy (Berwick and Weinstein, 1985) and the cervical smear (Grimes 1988) papers investigated a common problem. Both papers found that decisional uses of test information may greatly overlook the value that patients attach to the test. A similar evidence comes from a survey that attempted to establish the importance of factors beyond some medical definition of success in the provision of Assisted Reproductive Techniques (Ryan 1996 and 1998). Results of the studies suggest that there is some value in going through the service, even if couples leave it childless.

Returning to the classification of studies for the period 1984-1996, a fourth class includes three articles on assisted reproductive techniques that will be discussed later in the chapter (Neumann and Johannesson 1994; Granberg et al 1995; Ryan 1996), and a fifth class comprises two papers on WTP far autologous blood donation (Eastaugh, 1991; Lee et al., 1996). Finally, three papers form a special class of the contingent valuation literature in the health care field because they investigate general policies rather than specific medical procedures: Golan et al (1993) elicited WTP for supplementary insurance in the Israeli health care system, Eckerlund et al (1995) performed a contingent valuation study of the optimal size of the Swedish health care budget and Reis et al. (1990) investigated WTP of USA inner-city medical insured adults for a co-payment plan.

We were unable to classify the remaining seven studies published in the first period of the literature review. They concern paediatric confidential care (Fisher et al., 1985), hospital accommodation versus NHS nursing home accommodation (Donaldson, 1990), factors influencing the acceptance of hepatitis B vaccine by students (Pennie et al. 1991), bednet impregnation to protect against malaria (Mills et al 1994), a new antidepressant drug (O'Brien et al., 1995), repair of paediatric lacerations (Osmond, 1995) and satisfaction for ambulatory services (Ross et al., 1995).

A meaningful classification of the studies published between 1997 and 2005 according to the type of health problem investigated is even more difficult. Cardiovascular interventions are still frequent: one study referred to measures to reduce restenosis (Greenberg et al., 2004), one was on inhaled insulin for diabetic patients (Sadri et al., 2005), one estimated benefits of treating haemophilia (Carlsson et al., 2004), one estimated WTP for telemedicine services to patients with heart failure (Bradford et al., 2004) and a few studies investigated an hypothetical increase in heart interventions (Stewart et al. 2002; Luchini et al., 2003; Olsen et al., 2004a; Olsen et al., 2004b; Protiere et al., 2004; Ryan et al., 2004; Olsen et al., 2005). Twelve publications investigated screening/diagnostic services (Clarke, 2000; Liu et al., 2000; Papatheofanis, 2000; Wagner et al., 2001; Wagner et al., 2002; Thomas et al., 2000; Clarke, 2002; Forsythe et al., 2002; Takemura, 2005; Whynes et al., 2004; Whynes et al., 2005) but in contrast with previous papers the focus was not on eliciting WTP for non health related outcomes (e.g. the value of reassurance).

Nineteen papers derived from three major projects that deserve attention. Six papers reported results of the Eurowill project that was designed to assess the feasibility of the CV method as an instrument for investigating and measuring preferences of the general public about health care programmes (Stewart et al. 2002; Luchini et al., 2003; Ryan, 2004; Olsen et al., 2004a; Protiere et al., 2004; Olsen et al., 2005). The project investigated three health programmes: more heart operations, a new breast cancer treatment and helicopter ambulance service. These papers have severely challenged the validity of the contingent valuation method to offer guidance to policy making. Most of the work of the Eurowill project has underlined the inconsistencies of the answers that people provide to WTP questions.

The other project concerns the elicitation of WTP and its determinants for insecticide-treated bednets to prevent malaria in Nigeria (table 3.2) (Onwujekwe et al., 2001a; Onwujekwe et al., 2001b; Onwujekwe et al., 2002; Onwujekwe et al., 2003; Onwujekwe, 2004; Onwujekwe, Uzochukwu 2004; Onwujekwe et al., 2004; Onwujekwe et al., 2005a; Onwujekwe et al., 2005b). The papers derived from this project present a more optimistic picture of contingent evaluation. Overall, these ambitious and carefully designed surveys, that produced several methodological papers but at the same time gave valuable guidance to policy making, found the CV method reasonably valid and reliable.

Also the third major project was conducted in Africa and the WTP studies were part of a larger international cooperation project. However, in this study WTP was used to investigate community health insurance rather than a specific intervention. The study had a large sample size (almost 2,500 face-to-face interviews), allowed to make relevant methodological contributions (see above) and have been reported in four papers published in major international journals (Dong et al., 2003a, 2003b, 2004 and 2005).

Other two subjects of contingent valuation studies of this period are worth mentioning. Three papers (Pavlova, et al., 2004; Mataria et al., 2004; Asfaw and von Braun, 2004) investigated willingness-to-pay for quality improvement of health services in general. Instead of focussing on specific services, these studies assessed WTP for several dimensions of quality that could apply to a large variety of medical and public health services. Here, rather than providing elements to perform a specific cost-benefit analysis, the contingent valuation method allows to provide evidence of the people's willingness to invest for improving the quality of the health care system. The other subject that attracted our attention was vaccines (Liu et al., 2000; Morris & Hammit, 2001; Arana and Léon, 2002; Bishai et al., 2004; Suraradtecha et al., 2005; Liu et al. 2005;). Scholars have investigated hypothetical vaccines to prevent HIV/AIDS (Bishai et al., 2004; Suraratdecha et al. 2005), SARS (Liu et al., 2005), pneumonia (Morris & Hammit, 2001) and other diseases. It is dubious whether contingent valuation can really provide useful insights on programmes so hypothetical, although it is useful estimating potential return on investment of ambitious research programmes on vaccines. Anyway, these exercises look useful to investigate the attitude of respondents towards infectious diseases and towards hypothetical and existing prevention measures. In this respect WTP questions on a hypothetical vaccine may a part of a larger questionnaire to investigate attitudes and behaviour, and may be used to attract the attention of respondents.

3.3.6. Type of welfare measure and types of scenarios presented to respondents

As mentioned earlier, willingness-to-pay (WTP) is the maximum amount of money that individuals are willing to pay for a set of benefits, while willingness to accept (WTA) is the minimum amount of money that individuals are willing to accept as a compensation for a set of losses. Out of the 136 studies reviewed, only two papers clearly stated to use WTA to measure the value of a healthcare programme (table 3.3). The pilot study by Lindholm et al. (1994) assessed a community-based primary prevention programme against cardiovascular disease asking the following questions "Do you want the project (in which respondents were involved) to continue or do you want to stop, thereby reducing your annual tax by 300 (or 600/900) Swedish crowns?" In addition, as indicated by Gafni (1991), the study by Donaldson (1990) on continuing care for elderly persons used both WTP and WTA values to measure benefits and costs, respectively, of a change in allocation of NHS resources. However, Donaldson did not mention in his paper that he was using a WTA measure. More recently, Borisova and Goodman (2003) have explicitely measured WTA in terms of compensation for reducing travel time to get access to treatment.

WTP and WTA values can be measured in terms of compensating variations or equivalent variations. The former assumes that WTP/WTA are measured using the starting utility level, while the latter assumes that welfare changes ere measured according to the utility level after the decision being evaluated is implemented. Unfortunately, only a few papers openly reported whether compensating variation or equivalent variation was used. Johannesson et al. (1993) and Golan and Shecther (1993) made clear reference to a compensating variation measure, while Chestnut et al. (1996) and Stalhammar (1996) clearly stated that the WTP questions were asked in an equivalent-variation format. For the remaining papers we tried to identify from various elements what kind of measure was used. We reached the conclusion that the way the questions were framed and/or the context in which they were answered suggest that they all implicitly refer to compensating variations. However, this classification has to be considered with caution. More than the classification itself, the analysis of the literature on this specific aspect suggests that the distinction between compensation variation and equivalent variation is rarely made by researchers in the health care field and, probably, that such a distinction is not straightforward when conducting empirical studies. This seems to be confirmed by the fact that clear statements about the type of variations used are also largely lacking in papers written by economists and/or published in economic journals.

Table 3.3. Contingent valuation studies by type of welfare measure and type of scenario presented (1984-2005)

	First Author	Year of Pub.	Purpose of the study	Altruistic WTP Included?	Ex-ante or ex-post evaluation ?	Deterministic or probabilistic scenarios?
1	Thompson	1984	WTP, CV	No	Ex-post	Deterministic
2	Berwick	1985	WTP,CV	No	Ex-post	Deterministic
3	Fisher	1985	WTP, CV	No	Ex-ante & ex-post	Deterministic
4	Thompson	1986	WTP, CV	No	Ex-post	Deterministic
5	Grimes	1988	CV	No	Ex-post	Not applicable
6	Appel	1990	WTP, CV	No	Ex-post	Not applicable
7	Donaldson	1990	WTP/WTA, CV	No	Ex-post	Not applicable
8	Reis	1990	WTP	No	Ex-ante	Probabilistic
9	Eastaugh	1991	WTP, CV	No	Ex-post	Not applicable
10	Johannesson	1991	WTP, CV	No	Ex-post	Not applicable
11	Pennie	1991	WTP, CV	No	Not applicable	Not applicable
12	Johannesson	1992	WTP, CV	No	Ex-post	Not applicable
13	Golan	1993	WTP, CV	Not applicable	Not applicable	Not applicable
14	Johannesson	1993	WTP, CV	No	Ex-post	Not applicable
15	Johannesson	1993	WTP, CV	No	Ex-post	Not applicable
16	Lindholm	1994	WTP, CV	No	Ex-post	Not applicable
17	Mills	1994	WTP, CV	Not applicable	Not applicable	Not applicable
18	Neumann	1994	WTP, CV	Yes	Ex-ante & ex-post	Probabilistic
19	O'Brien	1994	WTP, CV	No	Ex-post	Probabilistic
20	Donaldson	1995	WTP, CV	No	Ex-post	Probabilistic
21	Eckerlund	1995	WTP, CV	Not applicable	Not applicable	Not applicable
22	Granberg	1995	WTP, CV	No	Ex-post	Probabilistic
23	Miedzybrodzka	1995	WTP, CV	No	Ex-post	Deterministic
24	O'Brien	1995	WTP, CV	No	Ex-post	Probabilistic
25	Osmond	1995	WTP, CV	No	Ex-post	Not applicable
26	Ross	1995	WTP	No	Ex-post	Not applicable
27	Chestnut	1996	WTP, CV	No	Ex-post	Deterministic
28	Kartman	1996	WTP, EV	No	Ex-post	Deterministic
29	Kartman (a)	1996	WTP, CV	No	Ex-post	Probabilistic
30	Ryan	1996	WTP, CV	No	Ex-post	Deterministic
31	Stalhammar	1996	WTP, CV	No	Ex-post	Deterministic
32	Weaver	1996	WTP, CV	No	Ex-post	Deterministic
33	Asenso-Okyere	1997	WTP,CV	No	Not applicable	Not applicable
34	Kartman	1997	WTP,EV	No	Ex-post	Deterministic
35	Lee	1997	WTP,EV	No	Ex-post	Deterministic
36	O'Conor	1997	WTP,CV	No	Ex-ante	Probabilistic
37	Olsen	1997	WTP,CV	Yes	Ex-ante	Probabilistic
38	Zethreaus	1997	WTP,EV	No	Ex-post	Deterministic
39	Donaldson	1998	WTP,CV	No	Ex-ante	Deterministic
40	Lee	1998	WTP,CV	No	Ex-ante	Probabilistic
41	Mathiyazhagan	1998	WTP,CV	No	Not applicable	Not applicable
42	O'Brien	1998	WTP/WTA,CV	No	Ex-ante & ex-post	Probabilistic
43	Onwujeckwe	1998	WTP,CV	No	Ex-ante	Deterministic
44	Ortega	1998	WTP,CV	No	Ex-ante & ex-post	Deterministic
45	Tambour	1998	WTP,EV	No	Ex-post	Deterministic
46	Chiu	1999	WTP/CV	No	Ex-post	Deterministic
47	Donaldson	1999	WTP, CV	Yes	Ex-ante	Deterministic
48	Dranitsaris	1999	WTP,CV	No	Ex-ante	Probabilistic
49	Matthews	1999	WTP/CV	No	Ex-post	Probabilistic
50	Onwujeckwe	1999	WTP,CV	Yes	Ex-ante	Probabilistic
51	Sorum	1999	WTP/CV	No	Ex-ante	Not applicable
52	Torrance	1999	WTP/CV	No	Ex-post	Not applicable
53	Cho-Min-Naing	2000	WTP,CV	No	Ex-ante & ex-post	Deterministic
54	Clarke	2000	WTP,CV	Yes	Ex-ante	Deterministic
55	Dranitsaris	2000	WTP/CV	No	Ex-ante	Deterministic
56	Estaugh	2000	WTP,CV	No	Ex-post	Deterministic
57	Liu	2000	WTP,CV	No	Ex-ante	Deterministic
58	Narbro	2000	WTP,CV	No	Ex-post	Deterministic
59	Onwujeckwe	2000	WTP,CV	No	Not applicable	Probabilistic

60	Papatheofanis	2000	WTP,CV	No	Ex-post	Probabilistic
61	Slothus	2000	WTP,CV	No	Ex-post	Deterministic
62	Suh	2000	WTP,CV	No	Ex-ante and ex-post	Probabilistic
63	Thomas	2000	WTP,EV	No	Ex-post	Probabilistic
64	Wagner	2000	WTP,CV	No	Ex-ante	Deterministic
65	Zarkin	2000	WTP,CV	Yes	Ex-post	Deterministic
66	Blumenschein	2001	WTP,CV	No	Ex-post	Deterministic
67	Dalmau-Matarrodona	2001	WTP,EV	No	Ex-post	Deterministic
68	Gyldmark	2001	WTP,CV	No	Ex-ante	Probabilistic
69	Morris	2001	WTP,CV	No	Ex-post	Probabilistic
70	Onwujekwe	2001	WTP,CV	No	Ex-post	Probabilistic
71	Onwujekwe	2001	WTP,CV	No	Ex-ante	Probabilistic
72	Wagner	2001	WTP/CV	No	Ex-ante	Deterministic
73	Arana	2002	WTP,CV	Yes	Ex-post	Probabilistic
74	Bhatia	2002	WTP,CV	No	Ex-post	Deterministic
75	Clarke	2002	WTP,CV	Yes	Ex-post	Probabilistic
76	Forsythe	2002	WTP,CV	No	Ex-ante	Probabilistic
77	Nocera	2002	WTP,CV	Yes	Ex-ante	Probabilistic
78	Onwujekwe	2002	WTP,CV	No	Ex-post	Deterministic
79	Stewart	2002	WTP,CV	Yes	Ex-ante	Probabilistic
80	Taylor	2002	WTP,CV	No	Ex-post	Probabilistic
81	Wagner	2002	WTP,CV	No	Ex-ante	Deterministic
82	Whittington	2002	WTP,CV	No	Ex-ante	Probabilistic
83	Zillich	2002	WTP,CV	No	Ex-post	Deterministic
84	Bhatia	2003	WTP,CV	No	Ex-post	Deterministic
85	Borisova	2003	WTP/WTA,CV	No	Not applicable	Deterministic
86	Dong	2003	WTP,CV	No	Not applicable	Not applicable
87	Dong	2003	WTP,CV	No	Not applicable	Not applicable
88	Foreit	2003	WTP,CV	No	Ex-post	Deterministic
89	Hammerschmidt	2003	WTP,CV	No	Ex-post	Deterministic
90	Luchini	2003	WTP,CV	Yes	Ex-ante	Probabilistic
91	Onwujekwue	2003	WTP,CV	No	Ex-ante	Probabilistic
92	Shiell	2003	WTP,CV	No	Ex-ante	Probabilistic
93	Tarasiuk	2003	WTP,CV	No	Ex-post	Probabilistic
94	Whynes	2003	WTP,CV	No	Ex-ante	Deterministic
95	Amin	2004	WTP,CV	No	Ex-ante	Deterministic
96	Asgary	2004	WTP,CV	No	Not applicable	Not applicable
97	Bissai	2004	WTP,CV	No	Ex-ante	Probabilistic
98	Bradford	2004	WTP,CV	No	Ex-post	Not applicable
99	Carlsson	2004	WTP,CV	Yes	Ex-ante	Probabilistic
100	Dong	2004	WTP,CV	No	Not applicable	Not applicable
101	Dranitsaris	2004	WTP/CV	No	Ex-ante	Probabilistic
102	Greenberg	2004	WTP,CV	No	Ex-post	Probabilistic
103	Hammerschmidt	2004	WTP,CV	No	Not applicable	Probabilistic
104	Mataria	2004	WTP,CV	No	Not applicable	Deterministic
105	Olsen	2004	WTP,CV	Yes	Ex-ante	Probabilistic
106	Olsen	2004	WTP,CV	Yes	Ex-ante	Probabilistic
107	Onwujekwe	2004	WTP,CV	No	Ex-ante	Probabilistic
108	Onwujekwe	2004	WTP,CV	Yes	Ex-ante	Probabilistic
109	Onwujekwe	2004	WTP,CV	No	Ex-ante	Probabilistic
110	Pavlova	2004	WTP,CV	No	Not applicable	Not applicable
111	Protiere	2004	WTP,CV	Yes	Ex-ante	Probabilistic
112	Rheingans	2004	WTP,CV	No	Ex-ante	Deterministic
113	Ryan	2004	WTP,CV	Yes	Ex-ante	Probabilistic
114	Ryan	2004	WTP,CV	No	Ex-post	Probabilistic
115	Whynes	2004	WTP,CV	No	Ex-ante	Probabilistic
116	Asfaw	2005	WTP,CV	No	Ex-ante	Not applicable
117	Barner	2005	WTP,CV	No	Ex-ante	Probabilistic
118	Bradford	2005	WTP,CV	No	Ex-ante	Not applicable
119	De Ridder	2005	WTP,CV	No	Ex-ante	Deterministic
120	Dong	2005	WTP,CV	No	Not applicable	Not applicable
121	Fautrel	2005	WTP,CV	Yes	Ex-post	Not applicable
122	Finkelstein	2005	WTP,CV	No	Ex-post	Deterministic

123	Hackl	2005	WTP,CV	Yes	Ex-ante	Deterministic
124	Hamelsky	2005	WTP,CV	No	Ex-post	Probabilistic
125	Ho	2005	WTP,CV	No	Ex-ante	Deterministic
126	Lee	2005	WTP,CV	No	Ex-ante	Deterministic
127	Liu	2005	WTP,CV	No	Ex-ante	Probabilistic
128	Masiye	2005	WTP,CV	No	Ex-ante	Deterministic
129	Olsen	2005	WTP,CV	Yes	Ex-ante	Probabilistic
130	Onwujekwe	2005	WTP,CV	No	Ex-ante	Probabilistic
131	Onwujekwe	2005	WTP,CV	Yes	Ex-ante	Probabilistic
132	Sadri	2005	WTP,CV	No	Ex-post	Deterministic
133	Smith	2005	WTP,CV	No	Ex-ante	Deterministic
134	Suraratdecha	2005	WTP,CV	No	Ex-ante	Probabilistic
135	Takemura	2005	WTP,CV	No	Ex-post	Not applicable
136	Whynes	2005	WTP,CV	No	Ex-ante	Probabilistic

Given the widespread use of public funding for health care services, externalities are expected to be of great importance. Actually, as many health services are private goods, the "added vale" of using contingent valuation to investigate WTP is strongly associated to the existence of caring externalities as they are not easily observable in real markets. A correct economic measure of WTP for health care services should provide the possibility to detect and measure these externalities. Despite this consideration, still few papers attempt to capture them. For the period 1984-1996 only the Neumann and Johannesson (1994) paper measured the altruistic component of WTP. They asked respondents about their WTP for a public programme that would have provided an IVF benefit for all Massachusetts residents needing treatment. This question was expected to capture altruistic externalities as most of respondents were not potential beneficiaries of the programme (see below for methods and results of this study). In the period 1997-2005 the number of studies designed to capture altruism is significantly higher: twenty out of one-hundred-and five (26%).

The demand for a large range of health services is such that people's welfare is increased by the financial risk protection provided by an insurance mechanism. Consequently, as Gafni (1991) claims WTP questions should be asked in the context of a hypothetical insurance purchasing. However, when we tried to classifiy studies according to this dimension we realised that only a limited number of them reflected an insurance-based perspective and that it would have been more relevant to consider whether respondents experienced the condition for which the intervention was suggested before the administration of the survey. According to this new classification, the large majority of studies used an ex-post perspective in that respondents who answered to CV questions were already in need of the service or were instructed to assume to be in that position. Roughly, about 50% of the studies present an ex-post perspective, but the use of an ex-ante perspective, often associated with hypothetical interventions, has become more frequent in recent years.

The last column of table 3 classifies articles according to the probabilistic or deterministic nature of the scenarios that were presented to respondents. Thirty-one studies (23%%) were not classified according to this criterion mainly because WTP was asked for programmes/interventions whose outcome of interest was not subject to uncertainty. Out of the remaining one-hundred-five studies, fifty studies (48%) provided deterministic scenarios and fifty-five (52%) allowed for some probabilistic elements in the description of the outcome. It appears that studies published more recently are more likely to provide respondents with probabilistic information.

3.3.7. Classification of the studies according to survey characteristics

Table 3.4 classifies studies according to the elicitation method used, the mode of data collection, sampling procedures followed, sample size and response rate. A variety of methods have been used to elicit WTP/WTA. One-hundred-seventy different evaluation methods were used in the 136 papers. Open-ended (OE) and payment card (PC) were reported in 43 (32%) and 39 (29%) papers, respectively. TIOLI was reported in 60 papers (44%) and the bidding game format in 28 (21%). Often open-ended questions were asked after another elicitation metod is used.

Open-ended questions were used in twelve studies out of thirty-two (37%) in the 1984-1996 period and only in six studies out of forty-two in the 2004-05 period (14%). Clearly, this method attracts little interest nowadays, probably because it does not resemble real market situations. It is worth remembering, however, that Onwujekwe and Uzochukwu (2004) argue that open-ended questions may resemble how donations occur.

Closed-ended questions have been becoming more popular in recent years: in 2000-05 forty-three papers (32%) report to have used this WTP elicitation method. The increasing interest in framing questions in a yes/no format probably reflects the recognition of respondents' difficulties in answering open-ended questions and the clear preference for this method expressed by the NOAA report (Arrow et al., 1993).

Nevertheless, the payment card format remains popular (25% of papers published since 2000) and economists are still debating on the merits of the different elicitation formats (Ryan et al., 2001; Cameron et al., 2002; Hanley et al., 2003; Smith, 2006).

Table 3.4. Contingent valuation studies by value elicitation method, mode of data collection and sampling characteristics (1984-2005).

	First Author	Year of Publ.	Value elicitation method	Mode of data collection	Sample and sampling method	Number respondents (% sample size)
1	Thompson	1984	Open-ended	Self-administ	Patients	36 (17%)
2	Berwick	1985	Open-ended	Face-to-face	Patients	62 (100%)
3	Fisher	1985	Payment card	Self-administ	Patients	165 (92%)
4	Thompson	1986	Open-ended	Face-to-face	Trial	237 (96%)
5	Grimes	1988	Payment card	Self-administ	Patients	84 (84%)
6	Appel	1990	Bidding game + open-ended	Face-to-face	Patients (random)	95 (95%)
7	Donaldson	1990	Bidding game	Face-to-face	Patients	107
8	Reis	1990	Open-ended	Face-to-face	Patients	149 (100%)
9	Eastaugh	1991	Open-ended	Face-to-face	Population	70 (100%)
10	Johannesson	1991	Open-ended + TIOLI	Mail	Patients	322 (66%)
11	Pennie	1991	Payment card	Self-administ	Benefic	435 (100%)
12	Johannesson	1992	Payment card	Face-to-face	Trial	61 (95%)
13	Golan	1993	Bidding game	Face-to-face	Population (random)	771 (75%)
14	Johannesson	1993	TIOLI	Self-administ	Patients	336 (65%)
15	Johannesson	1993	TIOLI + open-ended	Self-administ	Trial	692 (98%)
16	Lindholm	1994	TIOLI	Face-to-face+ mail	Patients	407 (77%)
17	Mills	1994	Open-ended	Face-to-face	Population	97 (100%)
18	Neumann	1994	Payment card	Self-administ	Population	231 (59%)
19	O'Brien	1994	Bidding game	Face-to-face	Patients	102 (78%)
20	Donaldson	1995	Payment card	Mail	Trial	260 (75%)
21	Eckerlund	1995	TIOLI	Mail	Population (random)	1021 (81%)
22	Granberg	1995	Open-ended	Self-administ	Patients	40 (85%)
23	Miedzybrodzka	1995	Open-ended	Self-administ	Patients	450
24	O'Brien	1995	Open-ended + TIOLI	Face-to-face	Patients	95 (76%)
25	Osmond	1995	Open-ended	Face-to-face	Patients	30 (100%)
26	Ross	1995	TIOLI	Face-to-face	Patients (random)	308 (76%)
27	Chestnut	1996	TIOLI +open-ended	Face-to-face	Patients	50 (100%)
28	Kartman	1996	TIOLI	Telephone	Patients	400 (87%)
29	Kartman	1996	TIOLI	Telephone	Patients	402 (92%)
30	Ryan	1996	Payment card	Mail	Patients	294 (42%)
31	Stalhammar	1996	Bidding game	Face-to-face	Patients	105 (100%)
32	Weaver	1996	TIOLI	Face-to-face	Population (quota samp.)	1263 (100%)
33	Asenso-Okyere	1997	Bidding game	Face-to-face	Population	50
34	Kartman	1997	Open-ended	Face-to-face	Patients	338
35	Lee	1997	TIOLI	Self-administ	Patients	235 (44%)
36	O'Conor	1997	TIOLI	Mail	Patients	148 (69%)
37	Olsen	1997	Payment card	Face-to-face	Population (random)	143
38	Zethreaus	1997	TIOLI	Face-to-face	Patients	104
39	Donaldson	1998	Open-ended	Face-to-face	Patients	150 (75%)
40	Lee	1998	TIOLI	Self-administ	Patients	412
41	Mathiyazhagan	1998	Open-ended	Face-to-face	Population (random)	918
42	O'Brien	1998	Bidding Game	Face-to-face	HMO	220

					enrolees	
43	Onwujekwe	1998	Bidding game	Face-to-face	Population (random)	1011
44	Ortega	1998	Open-ended	Face-to-face + telephone	Patients, population (random)	150
45	Tambour	1998	TIOLI	Self-administ	Patients	104
46	Chiu	1999	Payment card, open-ended	Face-to-face	Patient family members	174
47	Donaldson	1999	Open-ended	Face-to-face	Patients	150 (75%)
48	Dranitsaris	1999	Payment card	Telephone	Population	100
49	Matthews	1999	Payment card	Self-admininst	Patients, employees	41
50	Owujekwe	1999	Open-ended	Face-to-face	Population (random)	214
51	Sorum	1999	Open-ended	Mail	Parents	219 (68%)
52	Torrance	1999	Open-ended	Self-administ	Patients (trial)	240
53	Cho-Min-Naing	2000	Bidding game	Face-to face	Population	1480
54	Clarke	2000	TIOLI + follow-up	Telephone + mail	Target population (random)	372
55	Dranitsaris	2000	Open ended	Telephone	Population	80
56	Estaugh	2000	Bidding game	Mail	Patients	290 (77%)
57	Liu	2000	TIOLI + follow-up	Face-to-face	Population (random)	598
58	Narbro	2000	Payment card	Self-administered	Patients	374
59	Onwujekwe	2000	TIOLI + Bidding game	Face-to-face	Population (random)	719 (89%)
60	Papatheofanis	2000	TIOLI	Self-administ	Patients	87
61	Slothus	2000	TIOLI + follow-up	Face-to-face	Patients	179 (67%)
62	Suh	2000	Payment card	Self-administ	Patients	437 (72%)
63	Thomas	2000	Open-ended	Self-administ	Patients	1223 (70%)
64	Wagner	2000	Bidding game	Telephone	Population (random)	47
65	Zarkin	2000	Payment card	Face-to-face	Patients (quota)	393
66	Blumenschein	2001	TIOLI	Face-to-face	Patients	84
67	Dalmau-Matarrodona	2001	TIOLI	Face-to-face	Patients	228 (95%)
68	Gyldmark	2001	Payment card + Open-ended	Face-to-face	Population (random)	1349 (70%)
69	Morris	2001	TIOLI + follow-up	Telephone	Population (random)	1104 (76%)
70	Onwujekwe	2001	TIOLI + follow-up, bidding game	Face-to-face	Population (random)	709 (84%)
71	Onwujekwe	2001	TIOLI + follow-up, bidding game	Face-to-face	Population (random)	1908
72	Wagner	2001	Bidding game + open ended	Telephone	Target pop. (random)	1465
73	Arana	2002	TIOLI	Face-to-face	Population (random)	539 (77%)
74	Bhatia	2002	Bidding game + open ended	Face-to-face	Population (random)	1200 (100%)
75	Clarke	2002	TIOLI	Telephone	Target pop. (random)	372 (81%)
76	Forsythe	2002	Payment card	Face-to-face	Patients	519 (69%)
77	Nocera	2002	TIOLI, Dissonant Method, Payment Card	Telephone	Populatio (random)	1240
78	Onwujekwe	2002	Bidding game	Face-to-face	Population (random)	719 (89%)
79	Stewart	2002	Open-ended	Face-to-face	Population (random)	473 (45%)
80	Taylor	2002	Payment card	Self-administ	Patients	400 (89%)
81	Wagner	2002	Bidding game	Telephone	Population	1465

					(random)	
82	Whittington	2002	Payment card	Face-to-face	Population (quota)	234
83	Zillich	2002	TIOLI	Face-to-face	Patients	100
84	Bhatia	2003	Bidding game + open-ended	Face-to-face	Population (random)	298 (100%)
85	Borisova	2003	Open-ended	Self-administ	Patients	303 (66%)
86	Dong	2003	TIOLI + bidding game + open-ended	Face-to-face	Population (random)	1108 (86%)
87	Dong	2003	TIOLI + bidding game + open-ended	Face-to-face	Population (random)	2414 (90%)
88	Foreit	2003	Bidding games and open-ended	Face-to-face +self-administ	Various	About 13,000
89	Hammerschmidt	2003	TIOLI + Payment card	Face-to-face	Patients	92
90	Luchini	2003	Payment card	Face-to-face	Population (random)	163
91	Onwujekwue	2003	Bidding game, TIOLI with follow-up and structure haggling	Face-to-face	Population (random)	261 (66%), 267 (75%), 273 (58.6%)
92	Shiell	2003	Payment card	Face-to-face	Employees	112 (98%)
93	Tarasiuk	2003	Bidding game, open-ended	Telephone	Patients	252 (92%)
94	Whynes	2003	Open-ended, payment card	Self-administ	GP patients	2767 (40%)
95	Amin	2004	TIOLI with follow-up	Face-to-face	Population	324 (100%)
96	Asgary	2004	TIOLI, payment card	Face-to-face	Population (random)	2139 (86%)
97	Bissai	2004	TIOLI	Face-to-face	Population	1071 (70%)
98	Bradford	2004	TIOLI + follow-up	Face-to-face	Patients (trial)	126 (100%)
99	Carlsson	2004	TIOLI	Telephone	Population	609 (56%)
100	Dong	2004	TIOLI, bidding game	Face-to-face	Population (random)	2414 (90%)
101	Dranitsaris	2004	Open-ended	Telephone	Pharmacists, nurses	80
102	Greenberg	2004	TIOLI	Face-to-face	Patients (trial)	1642 (68%)
103	Hammerschmidt	2004	Payment card	Face-to face	Patients	92 (100%)
104	Mataria	2004	Payment card	Face-to-face	Patients (random)	499 (64%)
105	Olsen	2004	Payment card	Face-to-face	Population (random)	540 (100%)
106	Olsen	2004	Payment card	Face-to-face	Population (random)	168 (78%)
107	Onwujekwe	2004	TIOLI + follow-up, bidding game, structure haggling	Face-to-face	Population (random)	528 (59%)
108	Onwujekwe	2004	TIOLI + follow-up, open ended	Face-to-face	Population (random)	801 (89%)
109	Onwujekwe	2004	TIOLI + follow-up, bidding game, structure haggling	Face-to-face	Population (random)	528 (59%)
110	Pavlova	2004	Interval check list and open-ended	Face-to-face	Population (random)	990 (91%)
111	Protiere	2004	TIOLI + open-ended	Face-to-face	Population (random)	303 (95%)
112	Rheingans	2004	TIOLI	Face-to-face	Population	583
113	Ryan	2004	Payment card + TIOLI	Face-to-face	Population	578
114	Ryan	2004	TIOLI	Mail	Patients	325 (76%)
136	Whynes	2004	Payment card	Self-administ	Population	1401 + 202
115	Asfaw	2005	TIOLI + follow-up	Face-to-face	Population	550
116	Barner	2005	Payment Card	Self-administ	Patients	203 (41%)
117	Bradford	2005	TIOLI	Self-administ	Patients	366
118	De Ridder	2005	Payment card	Self-administ	Students	110 (96%)
119	Dong	2005	TIOLI + bidding game	Face-to-face	Population (random)	2414 (90%)
120	Fautrel	2005	Payment card	Telephone	Patients	119 (98%)

121	Finkelstein	2005	Payment card, probability statements	Internet-based	Population (panel quota)	1802 (100%)
122	Hackl	2005	TIOLI + follow-up	Face-to-face	Population	2536 (100%)
123	Hamelsky	2005	Payment card	Mailed	Patients	310 (65%)
124	Ho	2005	TIOLI + follow-up	Face-to-face	Patients	287 (100%)
125	Lee	2005	Open-ended	Telephone	Patients	517 (65%)
126	Liu	2005	TIOLI + follow-up	Various (mainly f-t-f)	Population (random)	1024 (77%), 488 (84%)
127	Masiye	2005	Payment card	Face-to-face	Population	274 (98%)
128	Olsen	2005	Open-ended	Face-to-face	Population	1240 (100%)
129	Onwujekwe	2005	TIOLI + follow-up, bidding game, structure haggling	Face-to-face	Population (random)	528 (89%)
130	Onwujekwe	2005	TIOLI + follow-up, bidding game, structure haggling	Face-to-face	Population (random)	321 (71%)
131	Sadri	2005	Payment card	Face-to-face	Patients	96 (92%)
132	Smith	2005	Open-ended	Face-to-face	Population	37 (94%)
133	Suraratdecha	2005	TIOLI	Face-to-face	Population	2524 (78%)
134	Takemura	2005	TIOLI	Self-administ	Population	152 (100%)
135	Whynes	2005	Payment card	Self-administ	Population	About 3000
136	Whynes	2005	Payment card	Self-administ	Population	1401 + 202

We identified five modes of data collection: self-administered questionnaires, mailed questionnaires, telephone interviews, face-to-face interviews and questionnaires administered thorugh internet. A large proportion of the surveys (87, 61%) were administered face-to-face, and mailed questionnaire (10, 7%) and internet-based questionnaire (1, 1%) were seldom used. In the 1984-1996 period self-administered questionnaires were used in 7 studies (22%), while questionnaires sent by mail were used in four studies (13%). These two modes of data collection present some common features as they require reading skills, they avoid interviewer biases and they are generally inexpensive. However, they tend to get different response rates: they are much higher in the self-administered questionnaires delivered by hand probably because respondents are probably more motivated to answer (it mainly depends on the way questionnaires are returned). Already in the 1984-1996 period, face-to-face interviews was the most common mode of data collection. It was used in nineteen studies (59%); in that period only two studies (6%) used telephone interviews.

Data from the 1997-2005 period show that face-to-face interviews are now the standard approach to conduct contingent valuation surveys: sixty-eight out of one-hundred-nine studies (64%) collected data through personal interviews, sometimes with the support of visual aids. This method, although expensive, has the advantage to maintain high the attention of respondents and to give the interviewer the possibility to keep under control the elicitation process. A few studies, mainly performed along clinical trials, still use self-administered questionnaires but this mode of administration is now less frequent. One paper was based on data collected in an internet-based survey conducted in the US (Finkelstein, 2005). This is the first internet survey we are aware of in the health care field and it has many features in common with the survey presented in the course of the following chapters.

Despite that sampling procedures are of great importance to generalise results and, hence, to offer useful material to policy makers, contingent valuation studies have not paid attention to this issue for long time. Generally, surveys conducted in the 1980s and in the 1990s were based on samples that were arbitrarily chosen from patients of a single health care institution. This probably reflected the methodological nature of many studies. However, testing the feasibility of using appropriate samples of the general population appears an important methodological issue too. In the 1984-1996 period only four studies elicited WTP from the general population and only two of them used a random sampling procedure. Interestingly, these studies did not investigate a specific medical treatment, but the size of the health care budget. Recent studies show a different picture in this respect. In the 1997-2005 period most of the studies are conducted with samples of the general population (56) and 23 of them report procedures to assure that they are representative of the adult general population living in defined geographical areas. Most of these studies used rigorous sampling procedures.

Sample size varies across the studies. In the period 1984-96 nine studies (29%) had a sample size of less than one hundred individuals. They generally collected information from patients using face-to-face interviews or self-administered questionnaires, and reached high response rates. Interestingly, 78% of these studies are not methodological and make conclusions of policy or clinical relevance. In the 1997-2005 peiod only nine studies (8%) were conducted with a sample size of less than 100 individuals and most of them were methodological.

During the older period fourteen (44%) studies were based on a sample size between 101 and 500 individuals. They were very heterogeneous in terms of mode of data collection, sampling method, type of study (methodological or decisional) and type of journal where they were published. In 1997-2005 50 studies (52%) have a sample size ranging from 101 to 500 individuals. In fact, while in the 1984-1996 period only 8 studies (25%) had a sample size of more than 500 individuals, in the period 2004-05 they were 42 (39%). In the latter period 25 articles reported surveys with sample size of more than 1,000 people. It is evident that there has been an improvement of contingent valuation studies over time: in more recent surveys samples are larger, sampling procedures are more accurate and face-to-face interviews are largely preferred over other modes of administration.

The last column of table 3.4 also presents the response rate of each study. The large majority of the studies have a high response rate, often above 80%. High response rates were high in the older studies mainly because they had a limited size and did not involve the general population. Most recent studies present similar high response rates, but these are obtained from larger samples and from members of the general population, presumably less motivated than patients surveyed on health care programmes of their interest.

"Poorly managed surveys can result in a multitude of problems such as falsified or incomplete survey forms, biased sampling, lost data, large numbers of people refusing to be interviewed or upset by interviewers and field assistants dropping out (i.e. the list goes on)" (Nyandieka et al., 2002). It appears that more recent studies, especially those conducted in developing countries, are better conducted. Also, recent articles report more information about important technicalities related to the surveys' conduction. Yet, however, the contingent valuation method is focused on theoretical issues and pays little attention to how the actual collection of data is managed. This is an area deserving future investment if the CV method wants to gain acceptance in the health policy community.

3.4. Contingent valuation studies on In-Vitro-Fertilisation and other Assisted Reproductive Techniques

A few studies have used the contingent valuation method to investigate In-Vitro-Fertilisation (IVF) or other Assisted Reproductive Techniques (ARTs). The first study was performed by Neumann and Johannesson (1994) and was carried out at the Harvard School of Public Health in 1992. The study was very ambitious and tried to address several critical issues. It attempted (i) to measure WTP for IVF treatments, (ii) to measure WTP to purchase insurance in order to have free access to IVF treatments, (iii) to measure WTP for a public IVF insurance programme and (iv) to investigate how people trade off the benefit of a public IVF program relative to a program which reduces the risk of mortality.

The contingent valuation study was based on 389 self-administered questionnaires distributed in classrooms, upon entering a conference or in mailboxes. Questionnaires were distributed to: 39 graduate students from the Harvard School of Public Health, 274 administrative officers from the Harvard School of Public Health and two health service centres at Harvard University, 48 nurses attending a conference and 28 physicians attending a seminar in Boston. The questionnaire used for the survey was nine pages long and clearly stated in the cover page that responses were anonymous and confidential. The first page was on IVF in general as it presented the way the procedure is administered and its main side effects. It also informed respondents that 10% of all American couples of childbearing are was infertile and that IVF was recommended for 5% of couples who seek fertility treatment.

In the first section the ex post perspective was explored: childbearing age individuals were asked if they would be willing to pay stated amounts for IVF in the event that they were infertile. Various assumptions about the probability that the procedure was successful were presented (chance of success equal to 0%, 25%, 50% and 100%) in separate questions. This section also reported that adoption was an available option and cost US $15,000.

Section 2 asked respondents if they would be willing to pay stated amounts for IVF insurance, assuming that they did not know their fertility status and that they had 10% chances to be infertile. This was the ex-ante perspective. Respondents were asked to state if they would be willing to pay for a one-time option to purchase lifetime insurance against IVF costs. Individuals who were not in childbearing age were told to skip this and the previous section.

Section 3 asked respondents if they would be willing to pay stated amounts in taxes for a public program consisting in providing free IVF services in Massachusetts. Absolute numbers (i.e. 300 expected beneficiaries of the programme) rather that percentages were used. Section 4 proposed a referendum on two alternative publicly funded programmes: providing IVF or reducing high-way fatalities. Respondents were asked to choose between an IVF programme resulting in 300 babies and a stated number of deaths avoided by implementing the other programme. Separate questions presented different assumptions about the effectiveness of the life saving programme (from 1 to 500 deaths avoided). Section 5 asked respondents to provide demographic information, an assessment of their (and their spouses) fertility status, current and expected household income and opinions about IVF.

The response rate of the survey was just under 60% (231 out of 389). Among 150 respondents who were potential child bearers, average ex-post WTP (respondents' WTP for the IVF in the event that they were infertile) was $17,730 for a 10% probability of having a child, $28,054 for a 25% probability, $43,576 for 50%, and 63,896 for a 100% probability. Ex-ante WTP (WTP for IVF insurance, assuming respondents do not know their fertility status) was $865 for a 10% probability of having a child, $1,055 for a 25% probability, $1,456 for 50%, and a $2,006 for a 100% probability. On average, respondents' WTP for a public programme with a 10% probability of success was $32. For higher probability of success average WTP was $38, $46, $62 for 25%, 50% and 100% probability of having a child, respectively.

Analysis of the results of the referendum to choose between IVF and a life-saving programme indicates that respondents evaluated that 300 additional IVF babies per year were equivalent to 35 prevented deaths. This means that, according to respondents' statements, 8.6 IVF babies were considered equivalent to 1 averted death. The implied WTP for a statistical life of the referendum (using the WTP for IVF) was $3.44 million.

Table 3.5 shows the marginal WTP per statistical baby that is the implied WTP to have a baby assuming that respondents knew to have a probability to be infertile equal to that of the overall American population. Obviously, for the ex-post WTP, the marginal WTP per statistical baby is equal to the product between WTP for the given chance that the treatment is effective and the inverse of this probability (e.g. WTP of $17.700 for a chance of 10$ results in marginal WTP equals to $ 17,700/0.1 = $ 177,000).

Table 3.5. Marginal willingness-to-pay per statistical baby.

	Probability of success (delivered baby following IVF)			
	10%	25%	50%	100%
Ex-post	$ 177,000	$ 68,827	$ 62,088	$ 40,640
Ex-ante	$ 1,730,000	$ 253,333	$ 328,000	$ 216,000
Public Programme	$ 980,000	$ 112,000	$ 100,800	$ 112,000

Source: Neumann and Johannesson, 1994.

Multiple regression analysis was used to test the influence of some demographic, clinical, social and attitudinal characteristics of respondents on the amounts that they were willing to pay and on results of the referendum questions. Coefficients of the regressions generally have the expected sign. Expected income, infertility status, inclination to use IVF and the desire to have more children were found to be positively associated with WTP (ex ante or ex post). The same variables were also negatively associated with the number of IVF babies per averted death.

The survey is presented by the authors as a pilot study to test the feasibility of the WTP method in the special context of IVF. The survey was feasible in the sense that about 60% of the sample answered to the question in a (presumably) meaningful way. However, sampled people, if compared with the general population, had better information about IVF, were better educated and were more motivated. Moreover, 77% of the sample were women and the mean age was 36.

Ex-post WTP for a statistical baby (WTP given that the respondent was told that he/she was infertile) was lower than the ex ante WTP (WTP given no information about the infertility status). With 10% probability of success at having a baby with IVF, the former was $ 177,000 and the latter was $1,130,000. This result is consistent with expected utility theory. As shown by Johannesson (1996) ex post WTP is a lower bound for ex ante WTP for an individual who is risk averse with respect to income. However, a ten fold difference between ex ante and ex post WTP for IVF appears too large. It appears exaggerated such a high risk aversion for the cost of IVF. In addition, it should be considered that in order to derive WTP values per statistical baby it is assumed that respondents understand that the expected probability of seeking IVF is 0.005. This value corresponds to the product of the probability that the couple is infertile and the conditional probability that IVF is recommended given that the couple is infertile. Both these probabilities are stated in the survey, but it is not clear whether respondents could really process these numbers.

Finally, table 3.5 shows that the ex ante WTP is higher than the WTP for a public programme. This conflicts with theoretical expectations. As WTP for a public program includes an altruistic component, it should be larger than WTP for an insurance programme. This inconsistency may be due to the difficulties that respondents may have found in answering questions involving risk evaluation and conditional probabilities.

In theory, results of this survey could have been compared to the costs of IVF treatments to measure welfare changes attributable to a programme providing this technology to infertile couples. However, Neumann and Johannesson (1994) did not make this comparison, probably because they were aware that survey results could not be generalised given how the sample was chosen. Instead, a straight comparison between WTP for IVF and treatment costs was reported by Granberg et al. (1995). These researchers from the University of Gotemberg conducted a contingent valuation survey with open-ended questions to elicit WTP for IVF among 47 couples seeking treatment in two clinics, one public and the other private. The couples were asked to state the maximum number of IVF treatments they would be prepared to undergo at different price levels. They were also asked their maximum WTP for having a child. The article does not report how the questions were administered. It seems likely that a self-administered questionnaire was used, even if the possibility that couples were interviewed cannot be excluded.

The paper provides few methodological details and results are presented in a very concise format. Response rate was 85% and WTP for a child ranged from UK £ 0 to £ 25,000. Mean and median WTP are not presented. It is only reported that 22 couples (54%) were willing to pay UK £ 10,000 or more. The article presents a figure illustrating a demand schedule for IVF from which it can be roughly estimated that mean WTP for a child was about UK £13,000. The study presents an estimate of the costs of the public IVF clinic and used this value to derive the cost per IVF delivered baby (dividing total annual costs by the number of babies delivered in one year). The cost per delivered baby was estimated at UK £ 9,410, just below the UK £10,000 WTP expressed by more than half of the couples. By comparing these two values the authors came to the conclusion that the benefit to the infertile couples was higher than the cost to the NHS. This is a poor quality study, mainly because the methods used were not well specified and the results were not clearly and fully presented. These deficiencies suggest avoiding comparisons between results obtained in this study and those obtained by Neumann and Johannesson (1994).

The third study is based on a survey mailed to 700 women attending a private infertility service in Sidney (Ryan, 1996). She investigated Assisted Reproduction Techniques (ARTs) from a different perspective to that adopted by the previous studies. She investigated whether factors beyond medical definition of success (a delivered baby) can explain WTP for ARTs. The survey investigated several factors which may be important to users in the provision of ARTs, including psychological outcomes, non-health outcomes and attributes of the process of treatment. Regarding psychological outcomes, Ryan was interested in feelings such as anxiety and stress provoked by the treatment or related to users' concerns about how the community views ARTs. She was also concerned with the feeling of regret. In the specific context of ARTs the feeling of regret refers to the fact that these interventions can be perceived as an infertile couple's last chance to have a baby of their own. If the opportunity is not taken, regret may be experienced later. As non-health outcomes she considered the utility that can be derived from information for its own sake, counselling and the provision of follow-up support. Concerning process attributes, she wanted to investigate the influence on WTP of factors as such as attitude of staff, waiting time, and continuity of contact with same staff.

For each of these factors she designed specific questions aimed at measuring their intensity as perceived by respondents. She used 0-10 rating scales (0 completely dissatisfied and 10 completely satisfied) or Likert scales from 1 to 5 (1 strongly disagree with a statement and 5 strongly agree). The questionnaire also asked respondents their maximum WTP for the treatment that they underwent or that they were undergoing. The WTP question used a payment card format where respondents were presented with a range of monetary amounts and were asked to circle the amount that they would be willing to pay. Finally, the questionnaire asked respondents to provide information about their age, education, children, household income, number of cycles of ART already attempted, whether or not they had left the service with a child, and perceptions of their chances of leaving the service with a child. Ordinary Least squares (OLS) regression models were used to identify the relative importance of attributes presented above, with WTP being the dependent variable and attributes hypothesised as significant predictors of WTP being the independent variables. A pilot study was carried out on 60 individuals who experienced ARTs to assess whether questions where understood and whether respondents were willing to answer the WTP questions.

The response rate to the survey was 42% and the item response rate to the WTP question was 96%. Twenty-nine percent of respondents had conceived on the programme. Respondents' WTP for an ART cycle ranged from AUS $ 425 to AUS $ 20,000 (2.1 AUS$ = 1 £). Mean and median WTP were AUS $2,506 and AUS $2,250, respectively. Questions looking at regret clearly indicated that there was some perceived benefit from going through the treatment. Of the respondents, 83% 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". Eighty-nine percent of respondents also agreed that even if they had left the service childless, they still would have been glad they had tried it.

Ryan also reports that respondents overestimated their chances of leaving the service with a child. Despite the fact that individuals attending the service were told that they had a chance of successful pregnancy between 15% and 25% (according to the type of ART used), over 50% of the sample agreed with the statement "when I started the programme I was very sure that I would leave it with a child".

Of the 339 individuals who answered the WTP questionnaire, 289 provided a full data set for the regression analysis. Results show that respondents who had a child from the programme had a WTP 30% higher than those who had not conceived yet or who had left the service childless. Further, the more respondents agreed with the statement "Even if I leave (or left) the service childless, I believe I will be (am) glad I tried it", the more they were willing to pay to undertake the service. However, all the other variables, including non health outcomes and process attributes, were found not to be significant predictors of WTP. Finally, as expected, individuals on higher incomes were more willing to pay for ARTs.

This study is a very interesting attempt to measure WTP beyond a narrow medical definition of success. She found some evidence supporting her hypothesis: results suggest that there is some value in going through the service, even if the couple leaves it childless, However process attributes and non-health outcomes were found to be un-correlated with WTP and the only question on psychological outcomes that was statistically significant may express the intensity of the desire of having a baby, rather than evidence in favour of regret theory. Moreover, the response rate to the questionnaire was relatively low and it is possible that respondents differed from non-respondents in many respects.

Later, Ryan conducted a similar study mailing a questionnaire to all clients undergoing IVF treatment at Aberdeen's Assisted Reproduction Unit (Ryan, 1997; Ryan, 1998). Again her main objective was to investigate the importance of psychological outcomes when going through ARTs, using the contingent valuation method (Ryan, 1998). Nevertheless, results of the WTP survey were also used by the author to make recommendations about funding ARTs, although the expression cost-benefit analysis was not openly used in the paper.

Although the study was similar to the previous one, its design differed in some important elements. First, a closed-ended (also labelled referendum or take-or-leave-it) approach was used to estimate WTP. This was in accordance with guidelines recommended in environment economics. Choosing this elicitation model requires to use logit models to identify the explanatory variables of WTP and to calculate central tendency estimates. Simply stated, within the framework of random utility theory, the probability that an individual will say "yes" to any given bid is estimated and then WTP is derived by integrating the estimated probability function (Ryan, 1998; see also chapter 4 of the thesis). Second, the survey included both ex-ante and ex-post evaluations. The majority of respondents were still going through the service when they were interviewed. According to Ryan's view they provided an ex-ante evaluation. Instead, a smaller fraction of respondents were not users of ARTs anymore as they either got a baby or they left the service childless. These patients provided an ex-post evaluation as they stated their WTP for a service they had already experienced. In addition, as the utility level of these respondents reflected their experience with ARTs, WTP measures are to be considered equivalent variations (they measure utility changes after the change). On the contrary, in the ex-ante evaluation WTP measures were compensating variations because they referred to the utility level before the use of ARTs. The ex-post evaluation also allowed comparing respondents experiencing positive and negative outcomes. Third, psychological outcomes were better specified through the use of two established scales to assess the psychological state of respondents: the Satisfaction with Life Scale (SWLS) and the Positive Affect Negative Affect Scale (PANAS). The use of these scales allowed better testing of Ryan's hypotheses about the value of psychological factors in the provision of ARTs. Particularly, she tested whether psychological feelings of "regret" and "disappointment" may explain the motivation to seek ARTs and whether these feelings may explain why the axioms of expected utility theory are violated. As mentioned earlier, regret theory is based on the premise that certain acts are taken to prevent regretting not having taken a given course of action later in life (Ryan, 1998). Disappointment is a psychological reaction to the results of an event (or a decision) not living up to its expectations. Results of the survey support both regret and disappointment arguments. WTP was substantial also for people who underwent IVF without success. Ryan argued that respondents felt there was a value in using IVF, even if they left the service childless, because of the feeling of regret. Support was lent to the importance of the feelings of disappointment by the association between WTP and the statement ""when our first attempt at IVF failed I was surprised". The more the respondent agreed with this statement, the higher the WTP. Upon this evidence Ryan argued that the psychological feeling of disappointment may be an important factor when looking at total utility from undertaking IVF.

In a different paper (Ryan, 1997) the author used the same data to elicit mean WTP for a cycle of IVF. Mean WTP was estimated to be in the British £ 5,000-5,100 range and evidence was provided of the internal and the theoretical validity of the contingent valuation method in the area of reproduction medicine. Estimated average WTP of users for ARTs was found substantially higher than current expenditure (£ 2,700; Ryan and Donaldson, 1995). Based on these results, the author suggests that the benefits of providing the service outweight the costs and that public provision of the service should be encouraged. It was not openly claimed that this was a cost benefit analysis, although both the terms costs and benefits were used in the paper.

3.5. Elicitation formats, use of internet and zero WTP in Environment Contingent Valuation studies

In this section we selectively review the contingent valuation literature outside the health care field to discuss specific issues of particular importance for the empirical study presented in the thesis. This part of the literature review complements what was presented in the previous sections and focuses on three major topics: a) the WTP elicitation format, b) the use of an electronic panel to administer the survey, c) the appropriate econometric techniques in case of zero and negative WTP.

3.5.1 Further considerations on the WTP elicitation format

In section 3.4 we reviewed the main WTP elicitation formats used in CV studies in health care. In this sub-section we complement our review with evidence produced by environmental studies, where the issue has been investigated more deeply. As presented earlier, four types of methods are used to directly ask WTP questions: open ended (OE), payment card (PC), take-it-or-leave-it (TIOLI) and bidding games. In more recent times, studies in the environment field have investigated a novel approach that is still rather uncommon for health care. In this approach respondents are provided a list of values and asked to decide among categories that describe the extent to which they may be willing to pay the stated amounts (Ready, 1996). Typically, respondents are asked if they are "definitely or surely" willing or not willing to pay defined amounts and are also let to opt for "do not know". This approach is often labelled as "multiple bounded" (MB) and measures stated preferences in terms of a distribution of WTP rather than of point estimates (Welsh and Poe, 1998). Such an approach increases the available information about preferences and can accommodate TIOLI as a special case.

Even after thousands of journal articles researchers still don't agree about the appropriate form of the valuation question (Whitehead, 2006). How to frame WTP questions remains a contested issue attracting several empirical studies with little consolidated evidence. Most comparisons (but not all) of TIOLI and OE methods performed in the environmental field suggest that TIOLI produce larger estimates (Brown et al., 1996). WTP estimates based on TIOLI questions are from 1.1 to 5 times higher than those based on OE. However, a few studies found TIOLI estimates lower and a review by Huang and Smith (1998) that used Monte Carlo simulations found that differences between the two methods are often due to misspecifications of the empirical model.

Other studies have compared TIOLI methods with payment card (PC). They have found that the TIOLI/PC ratio of the WTP estimates ranges from 2.7 to 4.4 (Cameron et al., 2002). Again, TIOLI generally produces much higher estimates and this is generally attributed to a "yes" bias (Kanninen, 1995; Holmes and Kramer, 1995).

In general, evidence from environmental economics, in line with that from health economics, shows that TIOLI estimates tend to be larger than those based on open answers or a range of values. Pair-wise comparisons are clearly useful but only signal convergence/divergence of methods. A more systematic approach to confront seven different value-elicitation methods was attempted by Cameron et al. (2002). The authors identified a common underlying indirect utility function for the identical good and compared different methods applied to different samples, pair-wise or pooled, across all samples in one unified model. They investigated 1) an actual dichotomous choice, 2) a first hypothetical dichotomous choice identical to the previous format but without the request to immediately pay for the good, 3) a second hypothetical dichotomous choice where different bids were presented to sub-samples, 4) an open-ended WTP question, 5) a payment card, 6) a multi-bounded discrete-choice questionnaire, 7) a conjoint analysis questionnaire.

Overall, this study made use of approximately 7,000 choices concerning environmental enhancement interventions and provides some important results deserving further investigation. Pooled-data models show that all methods but OE and PC appear to have a common underlying preference structure. Actually, findings simply show that there is no evidence to reject the hypothesis of identical utility-difference functions across the four "discrete" methods and thus further research in this vein is required. Nevertheless, these findings are interesting as they support the hypothesis initially formulated by Hanneman that the OE and PC methods might lead respondents to think about the WTP problem in rather different ways from that of discrete choice (Cameron et al., 2002).

3.5.2 The use of Internet for conducting CV surveys

In our empirical study we tested the use of internet to elicit WTP for funding a health care programme. This was the first attempt of this kind as, when we designed the survey, we were not aware of any attempt to use internet to elicit WTP in the health field. Even now, we are aware of only one study which used internet to administer a health care CV study (Finklestein, 2005).

Consequently, we turned to environmental studies and surveys in general to get guidance on the pros and cons of using internet. The aim of this section is to present the main issues related to the use of internet to administer CV surveys. The lure of this mode of administration is strong as the marginal cost of data collection is almost nil and the electronic format lends itself to easy data handling (Thurston, 2006). Sending some hundred questionnaires through internet saves time and postal costs (e.g. letters, envelopes, stamps). Motivating people to answer to internet survey may require incentives and specific investments (e.g. training), yet it appears that internet surveys tend to be much cheaper than those administered through other types of administration. In addition, data collection is much simpler and risk of data input and data computation are greatly reduced. Software manages the transfer of data from the electronic answer sheet to the databases, where the data is ready to be used by any statistical package.

In addition to these simple benefits (costs and data management), unlike with other modes of administration, Internet surveys make it possible to enhance the respondent's understanding of the programme in question. For example, it is possible to show drawings, photographs and graphs on the survey page; and it is also possible to provide links to other pages where additional information is available. The virtual nature of websites greatly expands the options to individualise surveys, for example by creating articulated pattern of links or by generating a large number of variants of the survey.

Web-surveys also make it possible to track the time spent by the respondents when pondering questions. Also, when links and other potential information sources are offered, it is possible to record what the respondent does so as to use time spent and sources accessed as potential explanatory variables. There are endless possibilities to take advantage of the technology (Thurston, 2006). The respondent can be forced in specific patterns (for example can be kept from looking ahead or back), virtual interviewers (for example of different races or gender) can be embedded in the page, and special aids and tutors can be used to facilitate answering the questions.

In front of the still unexplored wide range of benefits, internet surveys suffer from a major problem. Internet users are not representative of the general population. Only about 50% of the Italian population has some experience with internet and users and non-users tend to different in terms of age, educational status, income and, in Italy, even place of residence (urban versus rural and north versus south communities) (ISTAT, 2007). Whatever, the advantages of this new form of administration, internet surveys are constrained by the survey's sample and require adequate methods to improve representativeness.

In addition, a major problem with internet surveys is that we know very little about how they perform (Marta-Pedroso et al., 2007). We have found only one CV study that compared Web based to in-person interviews. As expected, the Web study had a much lower response rate (5.1% versus 84%). Instead, no significant difference was found about the susceptibility to information additivity and proportion of zero bids. However, findings of this survey conducted in Portugal indicate that Web based surveys generate more conservative estimates than personal interviews, whatever the payment vehicle used (taxation or donation). The authors concluded that the use of web based surveys is promising in the context of CV, despite the difficulties in drawing probability base representative samples (Marta-Pedroso, 2007).

Other studies, although they did not compare web based surveys to more traditional approaches, provide evidence of feasibility and show results that are similar to those obtained with other methods. For example, various national surveys in the USA tested the impact of information about global climate change and the Kyoto protocol (Li et al., 2004; Berrens et al., 2004). They found that WTP is associated with objective measures of respondents' effort to understand the issue and that WTP is sensible to variations of the Kyoto protocol that should make it more appealing to USA residents (Li et al., 2004).

There is a variety of types of Web surveys; they are summarised in Table 1 and can be classified into nonprobability and probability-based methods (Couper, 2001). Nonprobability methods recruit survey respondents without particular attention to representativeness of the population in question. Type 1 concerns Web surveys as entertainment. They cannot be considered a scientific survey and are mainly intended for entertainment purpose. These surveys can be launched by associations, newspapers and individuals just to collect answers on any type of issue. These surveys are generally used to attract attention on a topic and to activate discussions. In essence, they openly recognise that samples do not reflect any population and do not have any scientific meaning. Self-selected Web survey (type 2) are somehow similar as they freely recruit respondents on portals, frequently visited Web sites and dedicated sites. Given the recruitment procedures and the lack of any control of sample characteristics these surveys cannot reflect the characteristics of the population and should be considered very cautiously, even if conducted by prestigious institutions.

Table 3.6. Types of Web surveys (Couper, 2001)

Non-probability methods	Probability-Based Methods
1. Polls as entertainment	4. Intercept surveys
2. Unrestricted self-selected surveys	5. List-based samples
3. Volunteer opt-in panels	6. Web option in mixed-mode surveys
	7. Pre-recruited panels of Internet Users
	8. Pre-recruited panels of full population

The third type of Web Surveys concerns volunteer panels of internet users. Volunteers are recruited from well-travelled sites to form a database of potential respondents for later surveys. Panelists are typically recruited by invitation only and access to each survey is controlled through e-mail identifiers and passwords. Selection of panellists may be based on various criteria, including quota sampling and probability sampling methods. These methods make possible to control for the characteristics of respondents. Nevertheless, it should be kept in mind that the initial panel is a self-selected sample of volunteers.

The other methods listed in table 1 begin with probability samples of various forms (Couper, 2001). Probability sampling greatly improves representativeness of samples. Through knowledge of the population from which the sample is drawn and information on the process of recruitment, probability sampling permits measurement of non-response and thus can contribute to the improvement of survey design. In short, probability-based methods do not guarantee representativeness but provide valid tools to understand representativeness problems and to suggest improvements. There are two main approaches to achieving probability-based Web samples. One approach is to restrict the sample to individuals with Web access. The other approach is to start for a broader sample of the population so to try to include in the sample also individuals who do not have access to the Web.

Intercept surveys concern the first approach and target people visiting a particular Web site; they generally use systematic sampling to invite every nth visitor to answer to the survey. Clearly, the approach limits generalisation, but it can still be very useful to run costumer satisfaction surveys, evaluations and the like (Couper, 2001). A fifth approach focuses on specific populations for which web access is universal or quasi-universal. Here the limitation to those who are Internet users is not a problem. Employees of certain organizations, university students and members of scientific communities may constitute populations where Web access is normal. For such populations, a list of all individuals may be available so to use probability sampling to define the sample of interest. In this type of survey coverage is not a major issue; however, nonresponse rate remains a problem as internet surveys tend to have response rates lower that those of mail surveys of similar populations.

Web surveys can be combined with other modes of administration. Basically, following this approach researchers associate the Web survey to other methods to reach the part of the population that does not have access to the Web. This approach is popular in panel surveys of establishment (firms, schools, associations), where there is a long-term link between researchers and potential respondents (Couper, 2001). The seventh approach consists in pre-recruiting a panel of internet users and it similar to approach type 3. The main difference is that, in contrast with the earlier types where the panel was made of volunteers, in this type of survey panel members are recruited using probability sampling methods such as Randomized Digital Dialing. Through this approach samples should be constructed to be representative of those with access to the Web and it is possible to collect demographics and other data on Internet users and nonusers (and on those who did not accept to be panellists) to help understand the nature of coverage and nonresponse.

Finally, it is possible to design survey through probability sampling of full populations. This approach is similar to type 7 as it starts with a probability sample of the target population by using non-Internet approaches. However, while in the previous case the final sample was limited to Web users, in this type, researchers try to enrol non-Web users as well. Basically, using this approach, respondents are provided the necessary equipment and tools to become web users and, possibly, they are incentivised to regularly participate to surveys. Clearly, such an approach is costly because of high recruitment costs and thus makes economic sense only if it employs a panel design (several surveys regularly administered to the same sample). It is also the only approach that allows generalisation beyond current internet users and shows great promise for replacing probability-base surveys using more traditional methods (Cooper, 2001). One limitation of this approach concerns the auto-selection of panellists, in the sense that those who are recruited, even if they had no experience with the Web, are likely to be different from those who rejected the offer to be part of the panel. In addition, it cannot be excluded that becoming Web users can induce behavioural and attitudinal changes, thus introducing a new bias.

3.5.3 Appropriate econometric techniques in case of zero and negative WTP

Zero and negative WTP were expected in our study on a publicly funded program for IVF. We knew that some respondents were not in the market for private use, even in a hypothetical situation, and we suspected that some individuals could have had negative WTP for a publicly funded IVF programme. Nevertheless, we could not design a survey admitting negative values because it would have required splitting the sample according to some initial questions and this was unfeasible because of technical and economic constraints. In effect, the optimal strategy to investigate a service like IVF that can generate utility losses to some individuals for ethical reasons would be to ask an initial question to distinguish those who are in favour from those who are against and then to administer two separate sets of WTP questions.

This strategy was not feasible and thus we had to work on the large amount of zero WTP stated by the sample. If we limit our discussion to personal use of IVF and specific altruism for access to IVF there are three situations that need to be discussed. As far as personal use is concerned, there are individuals who have 0 WTP for IVF. These people are supposed to have a true 0 WTP as in the case of individuals that have no interest in a specific public good. They are not damaged by it, but they do not derive any benefit either and thus have true zero WTP. For this WTP question it does not make sense to investigate negative WTP because people who derive negative benefits from IVF simply state out of the market. The situation appears more complicated if a programme providing IVF with public funding is considered. In this case the good is a public good for some respondents and, likely, a public bad for others. In addition, it is also possible that people have a genuine 0 WTP in the sense that they attribute exact 0 value to such service.

To our knowledge, the issue of zero and negative WTP is unexplored in the field of health economics. Instead, in environmental economics it has attracted the attention of several scholars and some approaches to deal with the problem have been suggested. Starting with the issue of zero WTP, a possible approach in TIOLI Contingent Valuation studies is to use a spike model (Kristrom, 1997).

Let's imagine an individual facing a question to accept or reject a programme for a given sum of money A. Following Kristrom (1997) the project concerned the change of environmental quality from Z0 to Z1. Consequently it is possible to defined WTP for this change as

(1) $V(y - WTP, z^1) = V(y, z^0)$

Where V(y, z) is an individual's indirect utility function and y is income. Suppose that the individuals of a population evaluate the project differently. The probability that an individual's WTP is less than A is

(2) $prob(WTP \leq A) = Fwtp(A)$

Where Fwtp(A) is a right and non-decreasing function.

The expected value of WTP can be calculated by integration:

(3) $$E(WTP) = \int_0^{+\infty} 1 - F_{WTP}(A)dA \int_{-\infty}^{0} F_{WTP}(A)dA$$

According to Kristrom (1997), in the spike model the distribution function of WTP takes the following form:

$$
\text{(4) } Fwtp(A) = \begin{cases} 0 & \text{if } A < 0 \\ p & \text{if } A = 0 \\ Gwtp(A) & \text{if } A > 0 \end{cases}
$$

where p should be between 0 and 1 and Gwtp(A) is a continuous function such that Gwtp(0) = p and Gwtp(A) = 1 for A that tends to infinity

The previous equation shows the situation where a relevant faction of respondents have 0 WTP, which means that they are not in the market. The spike model can also be extended to the case where programmes generate both winners and losers. In the extended spike model the distribution of WTP can be written as follows:

$$
\text{(5) } Fwtp(A) = \begin{cases} Hwtp(A) & \text{if } A < 0 \\ p & \text{if } A \to 0- \\ p & \text{if } A \to 0+ \\ Gwtp(A) & \text{if } A > 0 \end{cases}
$$

where Hwtp(A) is a continuous and increasing function that has limit equal to zero for A that tends to infinity. Hwtp(A) refers to negative WTP, Gwtp(A) refers to positive WTP and p is the proportion of 0 WTP.

The models (4) and (5) can be estimated with various approaches. Model (4) can be estimated by using parametric maximum likelihood methods. It can also be estimated through a non parametric approach as suggested by Kristrom (1990). Similarly, model (5) can be estimated through maximum likelihood methods.

In the case of the extended spike model, negative WTP is addressed by formulating a questionnaire where both winners and losers can express their WTP for their preferred choice. In such a situation, researchers anticipate the presence of both winners and losers and prepare an adequate questionnaire.

How to handle negative WTP with TIOLI questions has been investigated extensively in the environmental field (Haab and McConnell, 1997; Bohara et al., 2001; Clinch and Murphy, 2001). One approach is to make use of Montecarlo simulations (Bohara et al., 2001). The study created a "true" distribution of WTP which included negative values and then it simulated a contingent valuation study. In short, investigators matched individual WTPi's to a particular randomly drawn bid to create a binary response variable Wi, where Wi is a Yes response if WTP ≥ bid, and Wi = 0 is a No response if WTP < bid. The study simulated 500 draws for different sample sizes (250, 500 and 1000 respondents) and made different assumptions concerning the underlying distribution when estimating WTP: normal (that allows for negative values), log-normal (that allows for positive values only), Weibull (that allows for positive values only) and two mixture models. The simulation was performed on the basis of three "true" distributions that varied only in terms of the number of negative bids (approximately 2%, 14% and 30%).

Results show that the assumption concerning the distribution is very important. In particular, it matters whether the chosen distribution allows for negative values if negative values are present. The normal distribution performs better than the other distributions as the log-normal and the Weibull inflate mean WTP. The Normal distribution also outperforms the Turnbull model, a popular nonparametric approach, if negative WTP is rather incidental. Restricting distributions to the positive domain significantly biases estimates, especially if the fraction of negative values is relevant. If negative WTP is rather frequent among respondents (e.g. more than 20%) neither a parametric approach (normal, log-normal, or Weibull distributions) nor the nonparametric Turnbull approach reasonably "solve" the problem if negative bids were not included.

Another approach, that can extend the solution to negative WTP when only positive observations are observed, is based on the analysis of censored data. Basically, in addition to assuming a particular distribution for WTP that allows negative values, it is assumed that the distribution is censored a zero. Observations at zero include negative observations because respondents could not express negative values. One way to deal with this case is to use a Tobit model or censored regression models (Woldridge, 2006). Open Ended questionnaires and payment cards generate WTP values that are normally censored at zero (Seung-Jun et al., 2001). If this is the case, OLS estimates are inconsistent. The Tobit model recognises the censoring problem but suffers from two limitations concerning the distribution of the errors. If heteroskedasticity occurs or there are errors not normally distributed, estimates are inconsistent. Therefore, the Tobit model may not solve the problems of OLS. A possible solution may be to use a censored regression model as that used by Seung-Jun et al. (2001). The authors assumed a Symmetrically-Trimmed Least Square (STLS) model. In a left (for example at 0) censored model some of the dependent variable yt are not observed if $y^* < 0$. This generates asymmetry in the distribution of the error terms. The basic idea of the STLS estimator "is to restore symmetry of the error distribution by symmetric trimming in such a way that the uncensored observations in the upper tails are replaced by their estimated symmetrically censored values". On one hand this approach addresses the problem of heteroskedasticity and non normality of errors; on the other, it also induces a loss of informative data concerning the upper tail of the distributions because of the trimming.

The issue of zero and negative WTP has not been investigated in the context of health care. Basically, health care is often a private good and people who do not derive any benefit from it simply decide to stay out of the market. However, for IVF the situation may be different because there may be individuals who suffer a welfare loss due to the use of the technique by others. We have claimed that this is a special case of negative caring externality. As already remarked, we could not design a survey that allowed negative WTP. However, we recognised the problem and looked for viable approaches in the literature. We found that in the environmental field the issues has been investigated and some approaches have been suggested. In the previous pages we selectively reviewed these approaches and found that, in order to investigate the role of true zero and negative values, we can work with spike or censored models.

3.6. The expected contribution of the present study

The study by Ryan appears to be a cost-benefit analysis because it compares the costs of providing a good to its value, measured in terms of WTP. ARTs users in Aberdeen attributed a value to IVF that outweighs cost. From their point of view the use of government resources to fund IVF produces a welfare gain. However, as Ryan (1997) recognises, an important question is whose values should be used in health care evaluations. As discussed earlier, Gafni (1991) argues that for a choice concerning government funding of health services it is the view of the community that is relevant. When valuing goods to be funded by the society at large it is important to capture whether people consider that those goods deserve to be funded collectively rather than privately. Instead, eliciting WTP from users neglects any consideration regarding the "social" merits of the goods and ignores the reasons that justify public funding.

Let's imagine that a group of people are invited to a free lunch and then that they are asked to state their WTP for the lunch. If we adopt a user perspective to elicit WTP we should conclude that the lunch has to be publicly funded if the average WTP exceeds its costs (assuming that cost actually refers to the opportunity cost of public funds). Basically, this is what Ryan (1997) did in her a cost-benefit analysis. Such an approach means to adopt a user perspective for goods that are private (in contrast to public goods that are non-rival and non-excludable).

However, for services private in nature, as in the case of most of health services, it may be important to include the perspective of non users for at least three main reasons. First, it is required in order to capture caring externality (Culyer, 1976) and option value (the utility an individual gains from knowing a service is there should they want to use it) (Weisbrod, 1964). It appears unlikely that the user of a service (e.g. IVF) makes an evaluation with reference to the option of having it in the future or to the utility derived by the use of others. Her evaluation is expected to be focused on the benefits derived from her immediate and personal use of the service. Second, in the case of private goods some members of the community (and thus potential payers of government funded programmes) can be non-users because they do not derive any utility from the service or may be even against the provision of such service. Therefore, surveys that elicit WTP from users depict the perspective of a biased sample of the community. By definition, users of a certain private good are not a representative sample of the entire community. Third, for private goods, given that they are produced and sold in actual markets, it may insufficient to state that community WTP outweigh cost to suggest public funding. Individuals should also agree with government funding, that is with transferring powers about the provision of goods to government decision-making.

A major problem with eliciting WTP from the community is that the community is unlikely to have good knowledge of health care interventions (Ryan, 1997). To inform individuals without experience with the good may be difficult and may result in information overload. Nevertheless, it appears unreasonable to exclude the perspective of non-users to evaluate private goods in order to decide whether they should be publicly funded. As cost-benefit analysis tries to offer guidance to public decision-making according to a societal perspective it appears appropriate to devise welfare measures from all types of members of society, including those who will never use IVF but may be required to pay.

The present study tries to go beyond the available evidence on public funding of IVF through a contingent valuation survey involving a representative sample of the Italian adult population. This sample allows eliciting WTP from non-users of IVF and designing welfare measures that reflect the perspective of all the categories of individuals directly or indirectly affected by IVF. In particular, the study can provide two main tests that are expected to inform the debate about whether ARTs should be publicly funded. The first test is partly consistent with the approach used in the studies presented above (Neumann and Johannesson, 1994; Ryan, 1996; Ryan, 1997; Ryan, 1998). In the first part of the survey, respondents are asked to imagine being infertile and being suggested the use IVF services. WTP elicited in this scenario reflects their ex-ante WTP for an IVF cycle and can be compared to the value obtained by Ryan. However, it should be borne in mind that mean WTP in this scenario results from all respondents, including those who would have not used IVF if they were infertile. Other things being equal, the mean WTP in the present study is expected to be lower that mean WTP obtained in the Aberdeen study (Ryan, 1997). By definition, the WTP of clients of an Assisted Reproduction Centre does not include the perspective of individuals who decide not to use ARTs services because of ethical arguments or because they attribute little value to these services. Despite the lack of information that the general population has on ARTs or other medical treatments, it appears more appropriate to measure welfare gains from a sample of the general population rather than from a selected sample of users. According to this test a cycle of IVF is worth being publicly funded if mean WTP outweighs its costs.

The second part of the survey attempts to elicit WTP for a publicly funded programme providing IVF to infertile couples. Here respondents are not required to imagine a hypothetical situation where they are infertile. They are asked to make a holistic evaluation of a programme that may be of direct interest only to some of them (those young or who are already seeking IVF services) but that it is funded through a tax increase. In this case the cost-benefit test is comparing total benefits, measured as the extrapolation to the national population of the elicited WTP values, to total costs, measured as the total costs of the IVF programme presented to respondents. Although, it may be challenging for respondents to process all the relevant information presented in the scenario, this approach appears the most consistent with an ex-ante, population based evaluation of net benefits of government funded programmes.

Chapter 4

Methods for measuring costs in healthcare

4.1. Introduction

This chapter deals with the costs of providing health services from a methodological and pragmatic point of view. It discusses how the cost analysis has been conducted in the economic evaluation literature and presents a few basic concepts and methods on how to perform sound analysis of costs. Two main conclusions stem from this chapter. First, there are various types of cost analyses depending on the decisions for which they are performed. Second, when an economic evaluation study is performed to decide whether a service should be publicly funded the most appropriate cost analysis method is full costing. This is the method used in this study to cost In Vitro Fertilisation Services. The aim of this chapter is to provide the background and the justification of this choice.

Cost analysis has not received particular attention in economic evaluation literature. Economic evaluators have generally preferred to focus on how to measure benefits. When they have turned their attention to costs, the main issue of discussion has been how to correct prices when they are not supposed to reflect social marginal costs. In the healthcare field cost analysis has not received great attention either. Major topics of methodological discussions in this field include measurement of benefits in CBA, discounting, sensitivity analysis and the use of appropriate statistical methods, identification and evaluation of outcomes, equity, and utility assessment. Concerning costs, methodological papers have mainly been focused on how to measure and evaluate indirect costs (leisure and working time lost, informal care, costs related to disability and premature mortality). So far, few authors have discussed methodological and practical issues concerning the cost analysis of goods and services (Luce and Elixhauser, 1990; Jacobs and Bachynsky, 1996; Graves et al., 2002; Wordsworth et al., 2005).

There are four main reasons suggesting that costing requires more attention in the economic evaluation of health care programmes.

i) Both cost-effectiveness and cost-benefit analyses compare costs to consequences. The incremental C/E ratio and the present value of net benefits are equally dependent on both the elements of the analysis; a 10% error in estimating costs has exactly the same impact of a 10% error in estimating benefits or effectiveness. Although effectiveness considerations may be prominent from the viewpoint of policy, economic results are equally sensitive to cost and economic data.

ii) Existing economic evaluation guidelines do not provide clear methodological indications or precise rules to evaluate costs. Although the US Public Health Service Panel on Cost-Effectiveness in Health and Medicine (Gold et al., 1996) and the second edition of the popular manual by Drummond and colleagues (1997) have made progress, a lot is still unknown and uncertain on how concretely to perform cost analysis.

iii) Economic evaluation papers generally use a few lines to explain how the cost analysis was performed and the authors often refer to unpublished material and personal communications.

iv) In many countries, including the USA, health care services do not have efficient prices because of relevant market imperfections and strict government regulation. As a consequence, it is rarely acceptable to refer to prices as proxies of opportunity costs.

Inaccurate costing may have relevant consequences. For example, in an Italian study (Fattore et al., 1997) the congruence between costs and reimbursement in an Italian NHS entity was checked. Using a full cost methodology (see below), the authors estimated the costs of 17 types of services provided by a Mental Community Service in Northern Italy. The results showed that the fee schedule adopted by the NHS to fund providers was not realistic. For example, for an outpatient psychiatric visit lasting between 20 and 30 minutes, the estimated cost was approximately 100,000 Italian Liras (€ 51.60). The official reimbursement fee for that type of visit was about 37,000 Italian Liras (€ 19.10). In this setting, using the reimbursement fee would have not been a good proxy of the opportunity cost. Nevertheless, the vast majority of economic evaluations performed in Italy use reimbursement fees only as a proxy for costs.

4.2. The concept of opportunity costs

In general, economic evaluation is based on the economic concept of opportunity costs. According to this principle, resources should be valued at an amount equal to their best alternative use. This concept is logically straightforward but relatively difficult to make operational.

Economic theory shows that if markets are perfectly competitive, then the price of products equals its opportunity costs. In such a situation cost analysis is relatively simple; prices of health services and other products are used as good proxies of opportunity costs. However, when markets do not exhibit the characteristics of a perfect competition, prices are no longer suitable guides to the opportunity costs of products. In short, these characteristics are the following: information symmetry between producers and consumers, the existence of perfect competitive markets for all products, the absence of externalities and public goods, and the absence of distorting incentives (Stiglitz, 1988). There is large agreement that these characteristics do not feature in the health care sector (Le Grand et al., 1992). In particular, information imbalances are relevant in both the market providing health care services and the insurance market.

In addition, in NHS-type systems, the prices of services are not generally available because provision is funded by tax-payers via the State, rather than by consumers. In these systems, opportunity costs cannot be measured through prices alone because the prices are not available. Recently, however, reforms in various NHS-type systems have introduced quasi-markets in the provision of health services. In quasi-market systems fee-schedules and other pseudo-prices are gaining ground. Italian hospitals (both private and public) are now partially funded through a fee-schedule system set at both national and regional level. The British NHS is planning to introduce a per case payment for hospital cases (called payment by result) and NHS trusts often have "price" lists for products purchased by Health Authorities or Primary Care Trusts.

Do fee-schedules and "price" lists used to regulate financial transactions within the public health care systems approximate opportunity costs? It is difficult to know the answer to this question, but the little evidence available to suggest that the answer is no. It seems unlikely that fees reflect opportunity costs because prices are not primarily set according to cost data. This is clearly what the Italian use of the Diagnosis Related Groups (DRGs) system shows (Fattore and Torbica, 2006). Ten years after their introduction, national and regional tariffs still do not derive from clear and transparent algorithms. The government performs costing exercises, but these exercises are very rough and tariffs appear subjected to minimally explicit and transparent rules. Tariffs are so important that policy makers, at least in Italy, do not intend to give up the opportunity to manoeuvre them with a high degree of freedom. Similar evidence between costs and tariffs can be found in other EU countries (Schreyogg et al., 2006) and was provided in a seminal paper written more than 20 years ago for the USA (Finkler, 1982).

In conclusion, very often prices cannot be used to costs services simply because they are not available in publicly funded healthcare systems. Also, when funding is related to activities as it is the case of the Italian National Health Service, charges rarely reflect costs because they are used to meet various policy objectives as well as to cover production costs. Finally, even when prices formed in the pure private market exist, as it is the case of IVF services, market failures (mainly information asymmetry) prevent prices to be reasonable proxies for the opportunity cost of resources.

4.3. The classification of costs in the economic evaluation literature

Traditionally, costs to be used in economic evaluation studies were classified into three general categories: direct, indirect and intangible (Luce and Elixhauser, 1990; Drummond et al., 1987). Direct costs concern the value of resources used in the provision of an intervention or in dealing with present and future consequences attributable to the intervention. Indirect costs on the other hand refer to resources whose use is attributable to any consequence of the intervention being evaluated and that cannot be counted as direct costs. Most indirect costs generally refer to productivity gains and losses attributable to illness or death. Finally, intangible costs refer to the monetary value of health losses per se (disutility generated by losses of health).

This classification has been criticised by the US Public Health Service Panel on Cost-Effectiveness in Health and Medicine (Gold et al., 1996) and by the Drummond et al. manual (1997). First, the distinction between direct and indirect costs does not appear to be clear. Is the time that an unpaid caregiver spends providing care a direct or an indirect cost? According to the definition presented above it is a direct cost; however, it is very common to find articles that report this type of cost as indirect. The term "direct cost" is not used consistently across the studies, which sometimes causes confusion (Drummond et al., 1997). Second, the term "indirect costs" is even more confusing as it is defined by difference (indirect costs are costs that are not direct). In practice, however, indirect costs generally refer to productivity gains or losses. As suggested by Gold et al. (1996), it is therefore more appropriate to label these costs as "productivity costs". Third, as noted by both Gold et al. (1996) and Drummond et al. (1997), the term indirect cost has another major interpretation. In management accounting an indirect cost is an item of cost that is associated by two or more cost objectives jointly but is not directly traced to each objective individually (Anthony et al., 1985). Fourth, intangible costs are not economic costs as they do not pertain to the use of resources. In addition, they are not strictly intangible because they can be measured and evaluated (e.g. using the WTP approach) (Drummond et al., 1997). It would be better if "intangible costs" are considered as benefits as they pertain to the contribution of health care goods to the utility function of individuals.

In this thesis we follow a straight approach as to what costs and benefit are. Benefits are monetary measures of changes of the utility function of individuals. The utility function of the individual depends on the present and future consumption of goods, given a set of characteristics of the individual. It may also include the consumption of goods by other people (caring externalities). Costs refer to the use of scarce resources that are the time of individuals, land, natural resources and goods/services used as means of production (land, natural resources and labour). According to this approach, the category of intangible costs does not exist because it pertains to the domain of benefits (a variation in people health concerns the utility function rather than an automatic use of scarce resources).

Focusing on costs, different classifications have been suggested. Drummond et al. (1997) classify costs according to three sectors: healthcare sector, patient and family, and other sectors. The US Panel (Gold et al., 1997, page 179) suggests keeping the term direct costs and to include in this category "the value of all the goods, services and other resources that are consumed in the provision of an intervention or in dealing with the side effects or other current and future consequences linked to it." As a consequence, direct costs encompass all types of resource use, including family, volunteer, or patient time. Concerning direct costs, the Panel also suggests to make a distinction between health care costs (tests, drugs, health personnel, etc.), non-health care costs (transportation, judiciary costs) and patient time costs (time spent by the patient seeking or undergoing an intervention). As far as productivity costs are concerned, the Panel identifies morbidity costs (costs associated with lost or impaired ability to work or to engage in leisure activities due to morbidity) and mortality costs (productivity losses due to premature death).

There are very thorny issues concerning costing. The most studied aspects concern how to measure working and leisure time spent by patients, productivity costs and mortality costs. The chapter discusses a different issue, which is how to cost health services when existing prices are not reasonable proxies of opportunity costs. In the next sections we present the type of cost information generally available in health care organisations and explain why they may or may not be used to cost services in cost-effectiveness and cost-benefit studies.

4.4. Accounting information and types of accounting

Decision making requires various kinds of information. Information is data (but also facts, perceptions, etc.) that improves knowledge. Information can be either qualitative or quantitative. The former type of information is expressed in terms of numbers, while the latter is discursive. Accounting information is a subset of quantitative information that is generated by an organisation (a company, a foundation, a local authority). There is no neat way to define accounting information. However, for sake of simplicity we can think of accounting as information expressed in monetary terms.

According to Anthony and colleagues (1985), accounting information can be classified in operating information, financial reporting and management accounting. Operating information concerns data on operations such as inventory, payroll, accounts receivable; they are the basis for the other two types of accounting information. Financial reporting requires operating information because financial statements are mainly derived by classifying and summarising this type of information. For all types of organisations (for-profit, private non profit and governmental)[1] the main purpose of financial accounting is to provide information to outside constituencies (Anthony and Young, 1988). Operating information is also the basis of management accounting, which is the function of providing information to plan, co-ordinate and control the organisation's activities (Anthony et al., 1985).

4.4.1. Financial reporting data and its limitations to cost health care services

Financial reporting and management accounting are radically different, although they represent areas that overlap. Most organisations (all private and many governmental) release financial reports using accrual accounting, that is a system where expenditures and revenues are adjusted to find the actual financial position of the organisation at the end of the accounting period.[2] For example, in income statements the cost of acquisition of fixed assets is not recorded; only a portion of them is reported (they are called depreciation expenses). Table 4.1 and table 4.2 present the balance sheet and the income statement of a hypothetical private infertility centre.

---

1 Anthony and Young (1988) suggest distinguishing between for profit and non-profit organisations. They then categorize non-profit organisations in governmental and private (tax exempt). While this classification may be adequate for the USA context and for the purpose of their book on management control, we find more useful here to classify organisations in three main categories: governmental (somehow controlled by the political system), non profit (private with general goals other than producing and distributing profits) and for profit.

2 Accrual accounting is focused on measuring the cost of resources consumed, rather than resources purchased. As a result this type of accounting makes use of accruals that are costs which have arisen but for which an invoice has not been received.

Table 4.1. Balance Sheet of Newborn Fertility Centre (As a December 31 2003 ('000 €).

Assets			Liabilities and Equity	
Cash..................................		400	Accounts payable..................	200
Supplies inventory...........		200	Wages payable...................	100
Equipment: Cost	4,000		Owed to bank.........................	4,000
Less: depreciation....200		3,800		
			Total liabilities.............	4,300
			Operating equity..................	100
Total assets..........................		4,400	Total liabilities and equity...	4,400

Table 4.2. Income Statement of the Newborn Infertility Centre 2003 ('000 €).

Revenues	
Patients revenue..	1,280
Interests...	50
Total revenues..	1,330
Expenses	
Wage...	450
Interest..	200
Rent..	150
Utilities..	50
Supplies...	130
Depreciation...	200
Other expenses...	50
Total Expenses..	1,230
Net income...	100

The balance sheet is the statement which contains the values of the assets and liabilities of an entity at a point in time. Table 4.1 reports that the Infertility Centre has assets for € 4.4 million and liabilities for € 4.3 million; the difference between assets and liabilities, 100,000 €, is its operating equity and corresponds to the net income recorded in the Income Statement. It should be noticed that both the statements present aggregate data. They provide useful information concerning the overall financial position of the company but cannot be used easily to cost products. Only in the case of a mono-product company would it be possible to calculate the unit cost of the product by dividing total expenses by the number of products. For example, if the company provided 1,000 IVF cycles it could be estimated that each cycle cost €1,230. However, organisations providing only one type of product are extremely rare. It is thus very unlikely that financial statements can provide useful information for economic evaluation studies.

Many governmental organisations still do not produce their financial statements using accrual accounting, although this type of accounting is becoming more popular in many countries. The main objective of financial accounting differs between private and governmental organisations. Outside parties in private organisations are mainly interested in having an overall picture of the financial position of the organisation, while in governmental organisations they are mainly interested in controlling the expenditure process.

Governmental financial accounting varies from one country to another and from one type of organisation to another. In Italy, NHS entities (Hospital Trusts and Health Authorities) presently use two types of accounting. Accrual accounting has just been introduced and (with important exceptions), tends to follow the accounting system of for profit organisations (regulated by the Italian Civil Law). The second type of accounting, "public accounting" was introduced when the NHS was established in 1978 and it is derived from the system in which all governmental organisations had to prepare their financial reports. In this type of accounting, expenditures are recorded at two different stages: i) when the entity becomes obligated to pay a certain amount of money and ii) when it actually pays it. Similarly, on the revenues side, amounts are recorded when the entity becomes entitled to receive that amount and when it actually receives them. Consequently, financial reporting provides two types of accounts: the debit/credit account where obligations and entitlements are recorded and the cash account where monetary transactions are recorded. Both the accounts appear in two documents: the “budget account”, prepared before the beginning of the accounting year, and the "final account" prepared after the accounting year has ended. Both documents are prepared under the responsibility of the General Manager, are approved by the referent political body (the Region for health care organisations in Italy) and are subjected to an accounting audit. The most important document is the "budget" as it authorises expenditure and allocates funds among different types of economic resources (personnel, goods, financial services). Various constraints regulate the way the "budget account" is prepared, the main one being that total expenditures cannot exceed total revenues.

Public accounting is a system designed to give political bodies control over the allocation of resources within the public sector. In contrast with private financial accounting, it is not an appropriate system to measure the economic performance of the organisation as it is less informative from an economic point of view. For example, it records investments for their entire amount and does not provide information on depreciation.

At least in Italy, public accounting suffers from severe shortcomings. First, "budget accounts" are estimates that can be changed over the accounting period (provided that a cumbersome procedure is followed) and "final accounts" are often prepared several months or years after the end of the accounting period. Therefore, either they do not fully reflect actual expenditures or they lack in timeliness. Second, single operations (e.g. wage payment of an employee, purchase of a type of pharmaceutical) are aggregated according to the nature of the resources acquired (personnel, pharmaceuticals) and not according to the area of intervention (prevention, hospital care, out-patient care) or to the organisational structure (hospital wards, administrative departments, etc.). Such a type of aggregation makes it difficult to extract information to cost products. Third, the system is not an accrual one; it produces information that does not reflect the value of the resources used in the accounting year. Two examples can explain this point. The system records all purchases of pharmaceuticals in a given period; whether the pharmaceuticals are used or not is impossible to know from the accounts. It may be the case that the organisation decides to purchase a large quantity of some items just to stock them. A high value is written in the books, but this does not reflect the use of that item in the accounting period. The second example concerns technology and equipment. in public accounting records there are two issues to bear in mind: i) the decision to purchase the equipment, ii) the corresponding payment/s. It is likely that the equipment is a long-lived asset that provides services for several periods after the expenditure to acquire it has been made. If in a particular year the organisation receives generous funding, its accounts will probably show high expenditures just because it used the capacity to spend. But these values have little to do with the use of capital equipment and, consequently, with the cost of providing health care services.

Both accrual (private) and non accrual (traditional public) accounting are not very useful to cost health services. This should not be surprising as they are intended to provide information to outside parties. In private organisations outside parties mainly consist of investors, creditors and providers. Their main concern is to know the overall financial position of the organisation so as to get guidance in their business operations. In governmental organisations financial accounting has a different main objective: to put allocation of resources under government control. In theory, governmental organisations should be interested in producing financial statements that justify the use of public money. This might include information on the costs of the main products delivered by the organisation. In practice, however, Italian NHS financial statements do not provide this data and are not generally useful to economic evaluators. Rather, it is management accounting that may prove useful to improve costing out products in the economic evaluation.

4.4.2. Types of management accounting

The purpose of management accounting is to provide information for managers in an organisation (Anthony and Young, 1988). While financial accounting tends to present an underlying unity targeted to outside constituencies, management accounting summarises information in different ways and for different purposes (Anthony et al., 1985). Traditional management accounting suggests three constructions that can be applied to costs: i) full accounting, ii) differential accounting and iii) responsibility accounting. Table 4.3 presents these types of accounting information and their use.

Table 4.3. Types of cost accounting information and their use (adapted from Anthony et al., 1985)

Cost construction	*Uses*	
	Historical data	*Future Estimates*
Full	Financial reporting (inventory and cost of sales) Analysis of economic performance Cost-type contracts	Programming Normal pricing decisions
Differential	NONE	Alternative choice decisions (e.g. make or buy, break even analysis)
Responsibility	Analysis of managers' performance	Budgeting

Source: adapted from Anthony et al., 1985.

4.4.3. Full cost accounting

Full cost accounting aims to estimate the total amount of resources consumed to produce a product or another object of interest. It can refer to goods (e.g. automobiles, books etc) or services (e.g. university courses, IVF cycles etc). It can also refer to different degrees of aggregation; the object of full costing can be the simple act of injecting a drug, the surgical operation to re-implant fertilised eggs, or the complete sequence of activities required to complete an IVF cycle. The unit of measurement chosen to be cost is called cost objective. The full cost of the cost objective is the sum of two components: i) the costs directly attributable to the objective (direct costs) and ii) a fair share of the costs incurred jointly in providing these and others products (indirect costs). Full cost accounting is used for various purposes; it provides estimates of the production of goods to be written in the income statement, and it may be the basis for contracts where buyers agree to pay the full cost of the product plus a negotiated profit margin. Estimates of future full costs can be used in long range planning (e.g. to estimate costs of a new hospital) and in normal pricing decisions, which is when prices are set according to costs disregarding market conditions. In addition, governmental and non-profit organisations as well as companies operating in regulated industries (e.g. transport, water, electricity etc) are often required to price their products starting from full costs.

In various sectors, many resource providers have published rules for full costing. In the USA there are rules set by the Department of Health and Human Services that apply to healthcare organisations providing care to patients under Medicare, Medicaid and other federal or state programmes. In England, prices agreed between NHS Trusts (defined in the contracts/agreements) and Health Authorities had to be set on the basis of costs as estimated according to the general criteria defined by the Department of Health (NHSME, 1993).[3]

As it aims to estimate the total amount of resources consumed to produce a product, full costing appears an appropriate methodology for costing health services. However, before discussing a few details of this methodology it may be useful to check if other methodologies may be appropriate.

4.4.4. Differential cost accounting

While the typical question of full cost accounting is "What did it cost?" differential accounting tries to answer the question "how will costs change under a proposed set of circumstances?" (Anthony and Young, 1985). The most common analyses using differential costs are: break-even analysis (estimating the volume of production at

[3] The circular provides minimum standards for costing across the NHS in order to improve contracting between purchasers and providers. These standards are based upon three basic financial principles: i) prices should be based on actual costs; ii) costs should be established on a full cost basis; iii) there should be no planned cross subsidisation between specialties, procedures or contracts.

which total revenues equal total costs) and make-or-buy analysis (the search for the cheapest option between managing production internally and contracting out). These types of analysis are very popular in private and governmental organisations. However, differential cost accounting focuses on given problems and thus cost estimates depend on the nature of the problem. As there is not a general way of labelling a given cost as differential or non-differential, differential cost accounting is a technique that varies according to the specificities of the situaions and does not make use of regular and systematic reports.

Anthony and colleagues (1985, 1988) state that differential cost accounting regards accounting information to be used when making alternative choice decisions. This definition is arguable, though. "Alternative choice decisions" is tautological as it is difficult to imagine a decision without an alternative. Without choices there are no decisions but only obligations. In addition, long-range planning and pricing are decisions as any plan or any decided price has at least one alternative. Differential cost accounting limits its analysis to costs directly implied by the choice. Therefore it needs not to refer to full costing of the alternatives; it focuses on costs that are emergent (new because of the decision) and ceasing (avoided because of the decision).

In general, differential accounting is not an appropriate cost methodology for cost-benefit analysis. However, if the new service to be evaluated requires a marginal change in the use of resources and no additional fixed costs, differential accounting may be more informative than full cost accounting. In such conditions, the production of the service would only imply an increase in variable costs and full costing would exaggerate the consumption of resources due to the production of the new service.

4.4.5. Responsibility cost accounting

In responsibility accounting organisations have a structure that defines roles, tasks and relationships among their members. This type of accounting refers to the organisational structure as it collects and reports accounting information about the use of resources and the output of organisational units. The units are generally called responsibility centres and are headed by a manager. Responsibility centres of a health care organisation may be hospital wards, the departments, the infertility centre, the payroll unit, etc.

Responsibility accounting interacts with full cost accounting. For example, in order to estimate the cost of products, full cost accounting makes use of information provided by responsibility centres. But the two types of accounting have different foci. Full cost accounting looks at products while responsibility accounting looks at the performance of responsibility centres. To a certain extent it can be stated that full cost accounting is just a technique while responsibility accounting is the heart of management control, which is one of the main management activities performed in an organisation.

4.5. Types of costs in management accounting

Generally, the economic evaluation literature refers to the book by Drummond et al. (1997) for categorizing costs. We report here a more sophisticated classification following a traditional textbook of cost accounting (Anthony et al., 1985). Table 4.4 reports a generic summary of cost types classified according to the three different accounting sub-systems presented above. Table 4.5 provides examples concerning infertility treatments.

Table 4.4. Summary of types of costs.

Full Cost Accounting	*Differential Accounting*	*Responsibility Accounting*
Direct: Costs traced to a single cost objective *Indirect*: Not-traced to a single cost objective; an equitable portion is allocated to the cost objective (overheads) *Full*: Direct costs + Indirect costs		Costs incurred in responsibility centres. *Controllable*: Manager can exercise significant (but not necessarily complete) influence *Non controllable*: Other costs, including committed and allocated costs
Capitalised: Asset to be amortised over several future periods *Product*: Direct + Indirect production cost of product *Period*: Expense of current period	*Variable*: Costs that vary proportionately with volume *Fixed*: Costs that do not vary with volume	*Engineered*: "Right" amount can be estimated *Discretionary*: Amount subject to manager's discretion; agreed on in budget process *Committed*: Will not change in the short run (a type of fixed cost)
Full costs are either historical costs or estimated future costs	*Differential costs are always estimated future costs*	*Responsibility costs are either historical costs or estimated future costs*

Source: Anthony et al., 1985.

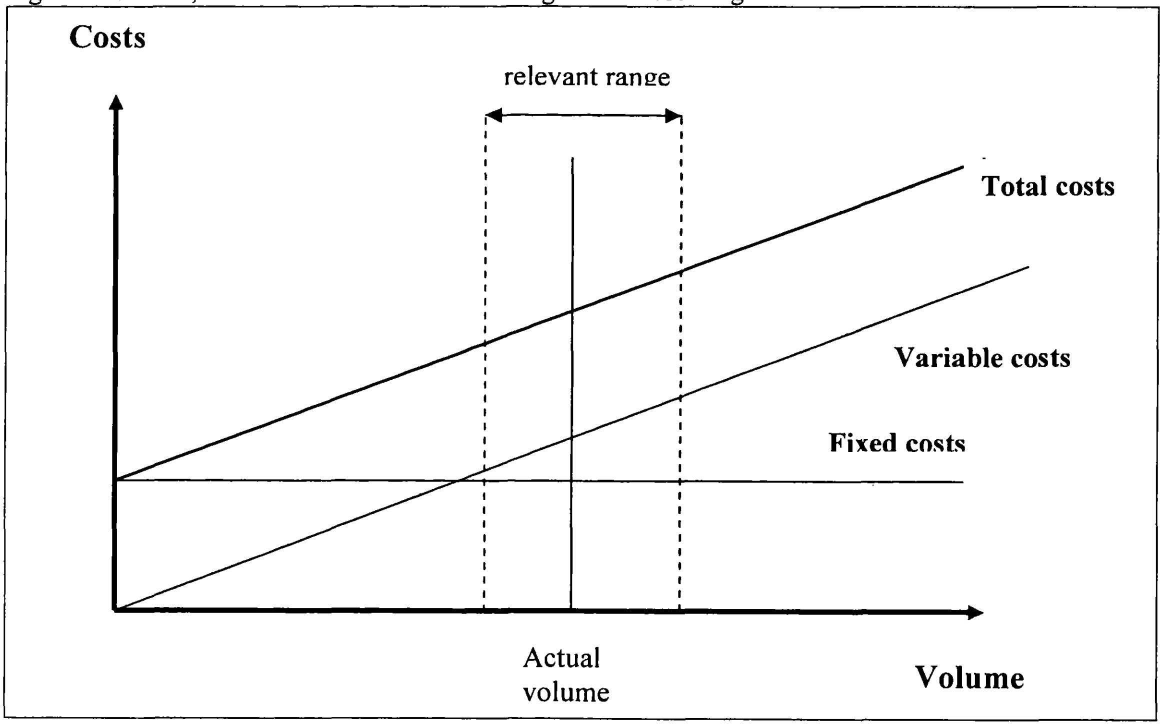
Table 4.5. Example of types of cost in an Infertility Centre belonging to an Italian NHS Trust (IVF cycle is the cost objective)

Full Cost Accounting	Differential Accounting	Responsibility Accounting
Direct: drug therapy Indirect: secretary services		Controllable: lab tests Non controllable: cleaning the Centre
Capitalised: most of the equipment of the operating room Product: the cost of producing an IVF cycle Period: the yearly cost incurred to keep an advertisement introducing fertility treatments in a popular website	Variable: the cost of the hormone therapy Fixed: the cost of cleaning the centre	Engineered: most of lab tests Discretionary: employees' participation to conferences and seminars Committed: depreciation on equipment and building

Full cost accounting makes use of two classifications. The first refers to the traceability to a cost objective; given a cost objective, a direct cost is a cost that can be traced to it. Assuming an IVF cycle is a cost objective, the surgeon time spent to re-implant retrieved eggs and all pharmacological treatments are direct costs. All administrative costs either of the NHS Trust or of the centre are indirect costs. It is not possible or feasible to trace them to IVF. The second classification is essential in financial accounting; capital expenditures are the costs of the acquisition, construction and installation of fixed assets. These costs are recorded in the balance sheet, but only depreciation is recorded in the income statement. This is because fixed (long-lived) assets make a contribution to production over more than one year and thus require having their costs apportioned over various accounting periods. Product cost is an expression mainly used in manufacturing companies and refers to the costs of producing goods. It therefore excludes marketing, selling, R&D and administrative costs. These costs are called period costs. It follows that i) the operating room and most of its equipment are capitalised costs, ii) under this classification, the "product" cost of IVF cycles exclude the cost of various activities not strictly associated with the act of producing, and iii) all administrative support, research activities and advertising (i.e. the production and distribution of a leaflet explaining IVF to the community) are period costs.

Differential cost accounting presents a classification very close to that used in economics, albeit with important caveats. It refers to the behaviour of costs with respect to volume (figure 4.1). An item of cost is variable when its amount varies proportionally with volume; an item of cost is fixed when whatever the volume the amount of resources remains unchanged. In addition, costs may vary according to volume more or less than proportionally and may follow different patterns (quadratic, logarithmic, etc.). The two caveats are important; cost behaviour always refers to a relevant range that is a range of volume where the behaviour described above is an acceptable approximation. The depreciation of the building where the Fertility Centre is located is a fixed cost, provided that the number of cycles delivered does not increase too much. For example, if the number of cycles more than doubles it will be likely that the space available does not suffice and additional space is needed. Either the increase in volume is impossible or it is required to acquire new space with consequent additional costs. The second caveat concerns time; in the long run many fixed costs are variable. When distinguishing between fixed and variable costs it is thus important to state the time period the analysis refers to. The cost of cleaning the Centre is fixed in the short term (e.g. one year), but becomes variable when the arrangement for its provision can be modified.

Figure 4.1. Fixed, variable and total costs in management accounting



Two classifications are useful in responsibility accounting. Costs can be either controllable or uncontrollable by the manager of the responsibility centre; controllability refers to the amount of resources used by the responsibility centre rather than to its unit costs. Lab tests used in an IVF cycle are controllable because, although provided by the Laboratory Department, are generally ordered by the Fertility Centre. Cleaning services are not controllable as it is very likely that the number of hours cleaners spend at the Centre are decided by the Cleaning

Department of the NHS Trust. According to the second classification costs can be engineered, discretionary and committed. Costs are labelled "engineered" when it is possible to establish a clear relationship between the amount of resources and the volume. Machine time required to perform a lab test is easily predictable; similarly, the lab tests required for an IVF cycle are predictable if clinicians follow protocols. Discretionary costs refer to resources that do not present a clear relationship with products. Their amount is generally decided by the head of the responsibility centre. The cost of participating to conferences is an example of discretionary cost. Finally, committed costs are those that are inevitable consequences of decisions previously made. If the Centre has stipulated a three-year contract with a famous clinician to monitor the quality of its activities, then the remuneration of the clinician is a committed cost for those three years.

4.6. The estimation of full costs

In one of the few papers reviewing costing methods in the economic evaluation of health care programmes Jacobs and Bachynsky (1996) argue that "improvements in costing can be made through the use of accepted cost-accounting techniques rather than through the continual reinvention of the wheel". They also state that proper costing techniques currently exist. However, as shown above, costs can be classified in various ways according to the purpose of the accounting system.

Responsibility accounting is nowadays very popular in many health care organizations. Under financial pressure, both public and private organizations are striving to keep costs under control through making organisational units more accountable for their use of resources. Consultants and other managers regularly receive reports on the costs of their units. Often, these reports are the basis upon which their performance is evaluated. However, these reports do not focus on products. Rather, they show actual costs compared with budgeted costs. Furthermore, controllability is a key element of this type of accounting. In order to make managers accountable, information has to focus on what they can control. The general administrative costs of the Health Authority or the costs of the activities required by the safety regulation are often disregarded in these reports as they concern aspects that are out of the scope of control of the vast majority of organisational units.

The manager of the Infertility centre controls the use of some of the resources presented in figure 4.2. Presumably, she has control over the amount of personnel time, the consumption of electricity, or the use of lab tests. However, she does not control the amount of cleaning services provided to the centre, or the costs of general administration. It is thus likely that the report he/she receives provides little information about these costs.

Figure 4.2. The step-down method.

	Departments	**Direct and indirect costs**	**Cleaning**	**Personnel**	**Laboratory**	**Total misison centre costs**
Support Centre	Clening	500				
	Personnel	100	30			
	Lab tests	500	30	20		
Mission centre	Internal Medicine	2.000	200	50	220	**2.470**
	O&G	3.000	200	50	200	**3.450**
	Infertility centre	500	40	10	130	**680**
		6.600	**500**	**130**	**550**	**6.600**

Source: adapted from Anthony and Young, 1985

Although responsibility accounting focuses on people and their organisational roles, it may provide useful information to cost products for economic evaluation studies. A well prepared cost report shows a detailed synopsis of the controllable costs of the responsibility centre. This piece of information may be an important step to estimate full or differential costs.

Full cost accounting is not widespread within Italian NHS care organisations. This is reasonable as NHS Trusts and HAs determine the prices only of a marginal part of their products. Full costing exercises are mainly performed in Italy by the national and the regional ministries of health to determine and revise tariffs (Fattore and Torbica, 2006; Adduce and Lorenzoni, 2004). On the contrary, British NHS Trusts appear to be keener to adopt full cost accounting, possibly because they are pressed into it by making sound contracts/agreements with Health Authorities.

Full cost accounting presents some key features. The first step of the procedures consists of assigning costs to the cost centre (the Infertility Centre).[4] According to Anthony and Young (1985) cost centres "...might be thought of as "buckets" into which an organisation's costs are classified and accumulated for purposes of full-cost analysis." Given the present situation of information systems in NHS Trusts and HAs, cost of amortisation of buildings and equipments, personnel, and utilities can be easily estimated and assigned to the cost centre. They are direct costs of the Infertility Centre as they are easily traceable to it (see above). Cleaning services and lab tests are also traceable to the Infertility Centre, although requiring more sophisticated information. If the NHS Health Authority has a responsibility accounting system, its reporting system will make available the number and the value of lab tests used by the Centre. If such a system is not in place, the number of lab tests can be estimated by surveying the clinicians working in the Centre. The value of clinical services used by the Infertility Centre can also be traced, either by taking advantage of responsibility accounting or by collecting information from the Cleaning Department. Lastly, figure 4.2 illustrates the fact that general administration, is clearly an essential component of the HA, and absorbs resources. Consequently, in order to estimate full costs of the Infertility Centre a fair share of general administration costs has to be allocated to it. The use of the word "fair" is not casual: there is no objective and true way to allocate these costs (Anthony and Young, 1988; Anthony et al., 1988; Drummond et al., 1997); a fair estimate means that the accounting system should try to find an allocation basis (i.e. the number of employees) that it is acceptable[5] and does not require excessive administrative costs.

The Infertility Centre provides various products; the list may include counselling to couples with infertility problems, stimulation therapies and more sophisticated types of infertility treatments as such as IVF. For the sake of simplicity, here we exclude the instances where the Centre provides products to other departments of the HA (the Obstetrics and Gynaecology (O&G) Department may require advise from the Infertility Centre) or to other health care organisations (the Centre may serve as

[4] According to traditional cost accounting, a cost centre is a cost objective for which costs are accumulated. Note that cost centres and responsibility centres may not perfectly overlap. In general, most responsibility centres are also cost centres. However, the reverse is not true; not all cost centres are responsibility centres. While the distinction between cost centres and responsibility centres may be relevant in many organisations, for sake of simplicity it may be disregarded in this chapter.

[5] The term acceptable may appear too vague. However, the level of precision of the allocation basis depends on how cost analysis is used to take decisions and the method, rather than being good or bad, it needs to be accepted by organisational actors to be useful.

semen bank for other centres). Despite this simplification, the Centre is not a mono-product organisation. It provides various products with different utilisation of resources. In order to estimate full costs of IVF cycles it is therefore necessary to quantify direct and indirect costs attributable to these specific cost objectives.

While building, personnel, equipment and utilities are traceable to the Infertility Centre, their attribution to products cannot be done directly; these items are indirect costs to IVF cycles. On the contrary, lab tests, although produced by another responsibility centre, can be classified as direct costs since each test can be attributed to a specific cost object.

Full cost accounting involves four major difficulties: choosing the allocation basis, choosing the allocation method, deciding between a process costing or a job order costing approach, and estimating the cost of using capital. We will turn to each of these issues in the following sub sections.

4.6.1. The allocation of indirect costs

Before discussing how to allocate indirect costs it is useful to introduce a basic distinction made in accounting for service organisations: cost centres can be classified in mission centres and support centres. The former (i.e. the Infertility Centre) refers to responsibility centres that contribute directly to the objective of the organisation, while the latter (i.e. the Personnel Department) contributes to the work of other responsibility centres, which may be either mission centres or other support centres.

Indirect costs are assigned to cost objectives by means of an overhead rate. Assuming that the Infertility Centre delivers only IVF cycles the overhead rate can be obtained by dividing total overhead costs by the number of cycles delivered. Since responsibility centres usually produce more than one type of product, output volume cannot be generally used to calculate the overhead rate. Intuitively, it is probably not acceptable to add the number of IVF and the number of counselling sections because they absorb different amounts of resources. In manufacturing companies the following volume measures are used as overhead allocation bases (Anthony et al., 1985): direct labour costs (the costs of the labour directly associated to the production of the product), direct labour-hours (the number of hours associated to the production of the product), machine-hours (the hours that the machine spent to produce the product) and prime costs (the sum of direct labour costs and direct material costs). Obviously, each measure results in different full costs and choosing the most appropriate allocation basis depends on the specific feature of production processes and the purpose of full costing. Given that personnel costs represent more than 50% of the overall costs in health care organisations, it appears reasonable to allocate overheads by the means of either direct labour costs or direct labour-hours.

Drummond et al. (1987) report various methods for allocating overhead costs. The simplest method is the direct method; it ignores interaction between support departments when any given support department's costs are allocated to departments providing services to final users. Coming back to the example of figure 4.2, this method means that the costs of the Cleaning Department are allocated to O&G, Internal Medicine, the Infertility Centre, but not to the Laboratory Department because it is a support centre. This method tends to underestimate costs of departments that make higher use of support department services.

The step-down method recognises that there is interaction among support departments. The Personnel Department provides administrative services to the Cleaning Department; but at the same time the Cleaning Department provides cleaning services to the Personnel Department (figure 4.2). In the step-down method a sequence of steps is chosen, usually by starting with the support department that provides the greatest service (in terms of costs) to the greatest number of other support departments (Horgren et al., 1996). The step-down process ends when overhead costs of the department that renders the least services to the least number of other departments are allocated. Drummond et al. (1997) mention other two methods which make full adjustment for interaction among support centres: the step down with iteration method where the step-down process is repeated various times to eliminate residual unallocated overhead costs, and the simultaneous allocation method where full adjustment is assured by solving a set of simultaneous linear equations.

4.6.2. Process costing and job order costing

There are two extreme approaches to product costing (Horgren et al., 1996): job-order costing and process costing. The process method is used when products are homogenous or when the aim of costing allows to treat products as homogenous. The most extreme way of using the process costing method is to consider the entire hospital as a mission centre; in its most simplified version it entails two steps to estimate the full cost per day: i) calculating all costs for the hospital, ii) dividing them by the number of days delivered. This approach may be useful when third-party payers reimburse a hospital according to an all-inclusive per diem amount (Anthony and Young, 1988) as was the case of private hospitals reimbursed by the Italian NHS before the 1992 reform (Fattore 1999; France et al., 2005). However, it is difficult to imagine an application of this method to the Infertility Centre for the heterogeneity of its products.

By contrast, with the job order method all direct costs are attributed to each specific product and an overhead rate is used to apportion indirect costs. Several Italian private hospitals use this method as a basis for pricing services delivered to private patients. This is also the prevalent method used by lawyers and other professionals to charge their clients. According to this method, for each IVF cycle major cost items are collected on a record, call job-cost record. For example, the record may collect health professional and equipment time spent on each cycle.

While the process cost method is too approximate, the job order method is probably too costly as it requires close monitoring of equipment and health professionals' time. It appears more suitable to follow a mixed approach which takes advantage of standard costing.[6] With a standard costing system, each mission centre defines the products it provides and the amount of direct labour, machine time and material for each treatment. With respect to the Infertility Centre, estimating standard costs entails: i) identifying the product objectives (inducement therapy, counselling, IVF cycles) ii) estimating the normal amount of labour time cost, equipment cost and lab test,[7] iii) adding one or more overhead rates.

---

[6] "A standard cost is a measure of how much a cost item should be, as contrasted with a record of how much it actually was" (Anthony et al., 1985, page 277)

[7] There are two types of standard costing: ideal standard and normal standard. Ideal standard costing implies maximum efficiency, while normal standard costing assumes a degree of efficiency that can be reasonably expected under prevailing conditions.

4.6.3. Capital costs

Surprisingly, the US Public Health Service on Cost-Effectiveness (Gold et al., 1996) does not discuss how to cost capital. In contrast, Drummond et al. (1997) present some basic material and provide practical suggestions. A thorough discussion of the theoretical problems associated with costing capital is beyond the scope of this chapter. Here, we simply identify some crucial and relevant basic issues.

First, capital costs are fixed costs and refer to expensive long-lived assets that contribute to the production of products/services over a period longer than one year. They can be either direct or indirect costs to cost objectives (cost centres or products). As the use of sophisticated and expensive technologies is increasing in health care, how to treat capital costs should be treated as a high priority in the research agenda.

Second, there are two main types of capital – land, building and equipment. Both the use of land and the production of building and equipment bring about opportunity costs as the land generally has alternative uses and thus the resources used to produce equipment and building. However, land, at least to a certain extent, lasts forever and without losing its characteristics while equipment has limited useful life. Therefore, equipment brings about an additional type of cost related to time and utilisation. In accounting, this type of cost is called depreciation. Almost all health care programmes require (or impact on) land and equipment. It is therefore essential that both "opportunity costs" and "depreciation costs" are properly measured.

Third, there is not a general rule to estimate depreciation costs because each piece of equipment has a specific (and uncertain) life span and its rate of utilisation, possibly impacting on the life span, is difficult to predict. In addition, obsolescence is also difficult to predict. Finally, even if the economic life span of the equipment is known, there is not just one way to apportion the economic value according to time. In financial accounting at least two general approaches can be followed – straight line depreciation and accelerated depreciation.

4.7. The role of full costing in the economic evaluation of healthcare programmes

One of the basic ideas of cost accounting is that the analysis is dependent on the decision that has to be made and on the time framework of interest. Let us assume that the Infertility Centre has launched a programme consisting of 10,000 leaflets explaining IVF treatments to be distributed to the community. The leaflets are already printed and the Centre has to decide whether to post them or to make them available in various public places. Does this decision require full costing? Certainly not; what matters here is the difference of costs of the two alternatives. Let us now assume that the Centre decided to post the leaflet but that a charity has already distributed similar leaflets to the same community. Probably, if the Centre had known the Charity's programme it would have not produced the leaflets. But does this mean that the programme is now too costly? Probably not; the cost of the production of the leaflets is a sunk cost, the decision has to be made only considering the cost of delivering the leaflets (and, obviously, the incremental benefit of distributing a second leaflet to the community).

Time framework is even more important. Let assume that the Centre increases production by 20%. It is likely that in the short-term only lab tests are variable, all other resources directly or indirectly involved in the production cannot be varied in the short-term. This means that the Centre and the support departments are required to increase productivity. However, it is very unlikely that an increase in production can be managed with an increase in productivity only. In the long term, unless there are permanent productivity gains, a stable increase in production must deal with an increase in the resources used. In the long run, even the number of employees of the personnel department is expected to be adjusted to the volume of activity of the Centre.

The US Public Health Service Panel explicitly recognises that economic evaluation (Cost-effectiveness Analysis) is an aid to decision-making. In this part of the chapter we have tried to show that organisations take different types of decisions and that each type tends to require different definitions of costs and different ways to analyse them. As a consequence, it is reasonable to think that there is no right way to perform an economic evaluation, but several possible ways depending on the specific issue of interest. For example, starting, increasing, decreasing or eliminating the provision of a service requires different types of cost analyses and results in different cost estimates. The aim of this discussion is not to prove that economic evaluation studies are useless in decision-making. Rather, it may help in pointing out that there is not "a right" way to do cost analysis and that there are various possible approaches that can be followed according to the specific type of decision that has to be taken.

There is a particular kind of decision that is attracting growing interest in the health care field. This is the decision to provide public coverage for particular treatments. Increasingly, health care systems are required to be more explicit about the services they provide under public funding. How much of the "basket" of services guaranteed to people is explicitly defined according to transparent and consistent rules varies from one country to another (Ham and Robert, 2003; Schreyogg et al., 2005). Nevertheless, some moves towards "explicit rationing" are made in a few countries and economic evaluation is increasingly seen as a possible aid to decision-making (Ham and Robert, 2003).

Decisions about public coverage of specific services have two main characteristics that are relevant from the point of view of the economic evaluation of healthcare programmes. First, they must refer to all organisations actually or potentially providing the service. Decision-making about policy is therefore expected to average out different situations. In other words, the decision to cover IVF with public funding should take into account that there are various types of centres (in terms of volume of services provided, degree of specialisation). Second, it seems reasonable to assume that policy-making is required to adopt a long-term perspective. Decisions concerning the exclusion/inclusion of services are expected to be long lasting and consistent over time.

These two characteristics suggest including all the costs that are variable in the long run. This is also the recommendation of the US Panel for the Reference Case. In effect, recommendation 12 of the cost analysis chapter states that "Costs in CEA should reflect the marginal or incremental resources consumed, rather than average costs, from a long-run perspective" and recommendation 17 states that "Variable costs, reflecting the value of those goods, services, and inputs that change because of the intervention being considered, should be included in the CEA, while fixed costs, which remain constant in the long run regardless of the level of production, should be excluded" (Gold et al., 1996: 209-210).

As shown above, full costing does not make use of the distinction between fixed and variable costs. The distinction is irrelevant because no costs are excluded by the analysis. Nevertheless, it should be feasible to exclude from full cost accounting certain types of costs. It is thus possible to use full cost accounting excluding long term fixed costs.

The economic evaluation literature does not provide practical rules to identify long-term fixed costs. Intuitively, however, in the long run almost all costs are variable. For example, administration, information technology equipment, space tend to be positively correlated with volume, although not always proportionally. The only exception we see concern costs related to the institutional arrangement of health care organisations. Whether the volume of production of the Health Authority doubles or halves there will be one and only one General Manager, Health Director and Administrative Director (to take the Italian case). In effect, these management positions exist and have similar costs in both small and large HAs and Hospital Trusts. But despite this type of costs, it seems acceptable to assume that virtually all costs tend to be variable in the long run and that full costs are an appropriate approximation of long run marginal costs.

Chapter 5

A cost-benefit analysis of In-Vitro-Fertilisation: Methodology and data

5.1. Introduction

This chapter presents the data and the methodology used to estimate benefits and costs of providing In-Vitro-Fertilisation (IVF) services in Italy. The first part of the chapter presents and discusses the contingent valuation survey used to estimate the monetary values that a sample of the Italian population attributes to the use of IVF services. The questions in the survey were framed to elicit two types of values of IVF services. People were first requested to give their willingness-to-pay for having access to IVF services in case they were infertile and wanted to have a baby. This part of the survey measured the use value of IVF, which are the benefits that respondents attribute to their personal use of the service. In the second part of the survey respondents were asked to state their WTP for a publicly funded programme providing IVF to Italian infertile couples desiring to have a baby. The questions of this part of the questionnaire elicit WTP for a public programme and thus are expected to catch the altruistic component of the programme, which is the WTP for benefits that are reaped by other members of the community.

The first section of chapter illustrates how the survey was built and conducted and explains the methodologies used to code and analyse the data. Section 5.2 focuses on the general characteristics of the survey; section 5.3 lists the socio-demographic information collected in the survey. Then a part of the chapter is devoted to the analysis of answers on knowledge and attitude towards infertility and Assisted Reproduction Technologies (ARTs) (sections 5.4 and 5.5 respectively). The central part of the chapter reports on the methods used to analyse the WTP data (sections 5.6 and 5.7). In addition to the analysis of the determinants of WTP for personal use and for the publicly funded programme, we present the models used to estimate mean and median WTP (section 5.9 and 5.10) and we detail the methods used to test the validity of the methods and to compare the elicitation formats that we used (sections 5.11 and 5.12).

Section 5.13 of the chapter presents the methods and the data used to estimate costs. It details how we obtained the full cost per IVF cycle that is used in the cost-effectiveness and cost-benefit analysis presented in the next chapter. The rationale for using a full costing methodology and details about the source and the type of cost data collected are discussed in chapter 4. The final section of the chapters shows how benefits and costs were combined to obtain a cost-benefit analysis that can inform policy decision-making.

5.2. General characteristics of the survey

The IVF questionnaire was sent through the internet to 6,435 Italian residents in late November 1999. These individuals agreed to be part of a panel created by CRA-Nielsen that is used for regular surveys on political issues and for marketing research. Participants on the panel received at their home computers one or two questionnaires each week in the form of an electronic file. They were requested to return the files with the answers within one week. Each participant of the Panel was equipped with a particular software programme which eases respondent's tasks.

We wanted to investigate WTP of a fairly representative sample of a national community. Methodological and practical considerations led us to choose an innovative method to interview individuals drawn from the general population: an internet-based survey. When the survey was designed very little literature was available on the use of internet to collect data from electronic panels. Even now little is known about this new method to administrate surveys, as very few studies have investigated its validity and reliability (Thurston, 2006).

Internet surveys have several advantages over other more traditional methods to administer surveys. In the context of the present survey they reach higher response rates, enforce question answering requirement, check errors and force respondents to follow predefined rules, allow for random question order, allow for automatic coding and data file creation, and ease the control of respondents' and sample characteristics (De Vaus, 2002). The main problem with internet-based surveys is that they allow for samples that are unrepresentative and that it is difficult to generalize from them. Given the limited and biased penetration of internet access, the choice of an adequate sample is critical. This survey used a so called "commercially recruited representative sample" (De Vaus, 2002), that is a panel of people who agreed to be regularly surveyed for money or in exchange of other benefits (e.g. free hardware and software). The survey was built according to a quota sampling procedure of the national population. From a list of families located according their home telephone number a sample of 5,227 individuals was randomly selected, in such a way that a minimum number of people meeting certain characteristics were included. These characteristics were gender, age (in 5 years brackets till 69 years), education (4 classes), employment status (5 classes), geographical area of residence (5 classes) and size of the municipality of residence (5 classes). The selection of individuals had to produce a minimum number of cases for each category to be included on the panel (even if not in a representative way). As a result the sample includes in a known way the characteristics above, so to make possible to estimate a weight for each member of the panel and thus to calculate statistical measures that are expected to represent the Italian population in respect of age, gender, education, employment status, geographical area and municipality size of respondent' residence.

This is by no means a random sample of the Italian population because the basic rule of random sampling that each member of the population should have the same (greater than 0) probability of being selected is clearly violated (Singleton and Straits, 1999). The sample was made of individuals who could get access to the URL where the survey was posted. There is no doubt that the use of personal computer, even if facilitated by another member of the family, may be not independent from the attitude towards IVF. However, the sampling procedure guarantees that i) an adequate number of respondents are included in each category, ii) the availability of information for the calculation of a weighting system to correct collected data. Overall, this sample can be considered fairly representative of the Italian adult population.

In addition, it is worth mentioning that the Italian ACNielsen/ISPO electronic panel is overseen by a prestigious research centre (ISPO) and it is regularly used by the most popular Italian newspaper (Corriere della Sera). Obviously, these facts do not prove that the sampling method is valid and appropriate for generalization; they only allow to state that the surveys obtained from this panel are deemed credible.

The choice of using a novel model of administration for the CV survey has various advantages as illustrated above. The main one is that we could reach a sample of the Italian population with a very limited budget. However, the use of a novel model of administration and the lack of literature and direct experience with internet surveys created a few constraints. The most relevant ones were due to the limited flexibility concerning the structure of the survey. We had limitations in terms of the number of questions and, more importantly, we had a very limited scope to work on sub-samples. Had we had more room for action we would have designed sub-samples to test various methodological issues. Instead, we could only design four variants to the survey and we used this opportunity to implement a TIOLI design.

Given that we also used a modified payment card approach to elicit WTP (see below), we could provide the original methodological contributions on the relations between the two elicitation methods, on starting bias and on assumptions about the probability distributions of WTP. Despite the impediments that we faced, our survey also tested relevant methodological issues on the use of CV surveys and provide evidence from a large sample of the general population. Our survey is one of the few attempts we are aware of that tests a healthcare CV survey on a rather representative sample of a national population. Most of the restrictions we faced should be seen in this context and may considered the price we had to pay to have a sample of the general population at affordable costs.

The survey includes three sets of WTP questions: two Modified Payment Card (MPC) questions, one for using IVF in case of infertility and the other for a publicly funded programme, and one TIOLI question. For each set we expected a large number o non-protest zero, which is a large number of respondents who have no WTP for the good under investigation. For the personal use of IVF this is because some people clearly think that they would not use IVF even if they would desire a baby. Especially in Italy, where the dominant Catholic Church is very influential and openly against IVF, people may see it as immoral, would not be in the market as a-priori decision and would not accept the good even at zero price. For a publicly funded programme the situation is more articulated. Two types of non-positive WTP are present. There are people who attribute a positive value to IVF but think that it should not be funded with public money. For example, they may argue that IVF satisfies a personal desire rather than a medical need (see for example the Dunning (1992) report in this respect). Other people could not derive any utility from IVF and would have negative WTP, in the sense of attributing a negative value to the fact that others use the good. They would suffer a loss in utility because of the use of a private good by other individuals. This is a special issue, rather unexplored in contingent valuation studies because, so far, negative WTP has been mainly investigated for public goods (Clinch and Murphy, 2001) and IVF is a private good and its externalities derive from a moral concern rather than from a direct loss of consumption opportunities.

The issue of zero WTP has not been widely discussed in the health economics literature, mainly because health services are often private goods and most studies have investigated the value perceived by the service user. The issue has been investigated in the environmental fields as many interventions (protecting a wild area or reducing pollution) have winners and losers, in the sense that interventions or goods, in addition to benefit some users and non-users, affect also people who directly suffer from the intervention (for example because they face limited opportunities to fish or hunt) (Clinch and Murphy, 2001; Bohara et al; 2001). This is not generally the case in health care. However, in environmental studies, it is the presence of negative effects of the public good to cause utility losses. For example, fishermen miss the pleasure of using a lake where fishing is banned and commuters suffer from longer journeys because of road-traffic constraints. Typically, in both cases some people have negative WTP because they directly suffer from the intervention. In our IVF case there is a similar situation in that there are people who suffer from the intervention in the sense that they suffer a utility loss if (other) people use IVF. To a certain extent this is a case of negative caring externality. In practice, it is important to distinguish these individuals from those who have no positive WTP but have nothing against IVF per se. We may argue that these people have 0 WTP but should not have a negative WTP for IVF because they do not suffer a utility loss from others' use of this technology.

Summing up, the case of IVF presents some peculiarities when measuring the WTP of a community. For the personal use of IVF there are people who would be in the market and other who would not be in the market even at zero price. For the publicly funded IVF programme we can reasonably assume that there are four possible categories of respondents: a) those who have zero WTP for the programme (for direct interest and for altruistic reasons); b) those who have 0 WTP for personal use, in the sense they would never use IVF in case of infertility, but have positive IVF for a public programme; these people present a special case of altruism attitude that we label here pure altruism as they are willing to pay for a service that they would not use even in case of need; c) those who do not have moral arguments about IVF but think that it should not be publicly funded (these people should have zero WTP for the public programme), d) those who are "morally" against IVF and have negative WTP in the sense that they suffer from the fact that other members of the community uses it (they procreate new members of the community using this technique); these people have negative WTP as they would like to pay not to have an IVF publicly funded programme.

In the design of the questionnaire we did not introduce the possibility that respondents have negative WTP for a public programme providing IVF to infertile couples. This would have required a completely different questionnaire design and would have greatly complicated our electronic survey. Nevertheless, we can make various assumptions about the error distribution and censoring so to allow that patients may have had a negative WTP for the programme even if we did not openly design this possibility (Hanemann and Kristrom, 1995; Kristrom, 1997; Seng-Hoon and Seng-Jun, 2002). In addition we can detect respondents who have pure altruism as we have WTP data about both personal use of IVF and funding a public programme.

5.3. Socio-economic questions

The following individual demographic, social and economic information was collected in the survey:

- age
- gender
- education (4 classes: less than 8 years of education, between 8 and 13 years, high school diploma and university degree)
- employment status (7 classes)
- municipality of residence
- marital status
- number of children
- Self-reported socio-economic status (5 classes: low, low-middle, middle, middle-upper, upper)

- Monthly family income after taxes (13 classes from less than approximately 400 Euros per month to more than approximately 4,000 Euros per month).[8]

The questionnaire included two specific questions on the economic condition of respondents: self-reported socio-economic status and self-reported household income. In the first question individuals were asked to attribute their families to five social groups (from lower to upper ranks). In the second question, individuals were asked to reveal their family monthly income after taxes choosing between 13 brackets. The question explicitly included a "no answer" option.

5.4. Respondents' knowledge of infertility and Assisted Reproductive Techniques (ARTs)

Four questions were designed to collect information on respondents' knowledge about infertility and Assisted Reproductive Techniques (ARTs). The first question was about respondents' acquaintance with couples who desire to have babies and are unable to do so naturally (table 5.1). The second question investigated whether respondents knew IVF and from what sources they had obtained information pertaining to it. Then, respondents were asked to rate their knowledge about IVF on a five-level scale. Finally, respondents were asked if they had met someone who used IVF or other ARTs.

Table 5.1. Questions about respondents' knowledge about infertility and Assisted Reproductive Techniques

Are you acquainted with any couples who desire to have babies and cannot have them? Yes No
Did you know about In-Vitro-Fertilisation before? No, It is the first time I heard of it Yes, I have heard of it from the TV and/or newspapers Yes, I have heard of it from relatives/friends Yes, I have got informed through specialised readings Yes, I have got informed from the doctor
How would you rate your knowledge on In-Vitro-Fertilisation? Very limited Relatively limited Sufficient Relatively wide Very wide
Have you ever met someone who used IVF or other ARTs? Yes No

[8] Appendix A reports the questionnaire and the English traslation

We ran regression analyses to identify the independent variables that explain the variability in the answers to the four questions presented above. The statistical models included as potential explanatory variables the following characteristics of the respondent:

- age (discrete variable ranging from 18 to 69, the age limits of the study);
- fertility age (yes (1) for individual between 18 and 45, no (0) otherwise);
- gender (female = 1, male = 0);
- education (4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4));
- employed (yes=1, no=0; retired people, students and housewives/husbands included in the non employed group)[9];
- geographical area of residence (3 groups: northern, central and southern - including isles- regions modelled as two dummy variables);
- residence in a city which is "province capital" (county town)[10] (yes=1, no=0);
- marital status (1 if married of living with a partner and 0 otherwise);
- Number of children (discrete variable);
- Socio-economic status (5 classes ordered from 1 to 5 with 5 as the highest social economic status);
- Monthly family income after taxes (13 values derived by the classes presented in the questionnaire)[11].

Three assumptions were made to generate income values from the 13 income classes presented in the survey. First, responses indicating classes with defined lower and upper limits (e.g. between It Liras 800,000 [€ 413] and It Liras 1,000,000 [€516]) were set at the mid value of the range (in this case It Liras 900,000 [€465]). Second, for the lowest income class (less than It Liras 800.000 [€ 413]) it was assumed a value of It Liras 400,000 [€ 207] (half the value of the upper limit). Third, for the highest income class (more than It Liras 8 Million [€ 4,132]) the value was set

---

[9] In the empirical study one male respondent stated that he was a "casalinga" which is the Italian expression to indicate a person who is fully dedicated to managing the house.

[10] Italy has four political tiers directly elected: the State, 20 Regions, 103 provinces and about 8,000 municipalities. All major Italian cities are province capitals and the smallest ones have about 20,000/30,000 inhabitants.

[11] See Appendix A for the exact wording of the question on family income.

at It Liras 16 Million [€ 8,263] (twice the value of the lower limit). These assumptions made it possible to use cardinal values in several statistical models presented in the thesis.

Most of the independent variables are dichotomous or categorical (without any logical order) and were thus transformed into dummy variables. For age, income and number of children, that are discrete cardinal variables, we checked normality and assessed possible transformations that can improve goodness of fit (R^2, Likelihood Ratio and prediction value). We also visually checked normality of socio-economic status and level of education that present ordered values.

For each dependent variable three dataset were used: i) the "full dataset" including all possible independent variables presented above (dataset a), ii) the "full dataset" including the variables selected by a backward stepwise regression (dataset a1), iii) a model based on the "income dataset" including the variables selected by a backward stepwise regression (dataset b). In the "full dataset" all the 5,739 observations registered in the empirical study were included, while in the "income dataset" the 933 respondents who preferred not to state their income were recorded as missing values thus resulting in 4,806 useful observations. In regressions a1 and b the stepwise procedure starts from the model with all the independent variables and then, step-by-step, it removes the independent variables having the least absolute partial correlation with the dependent variable, controlling for the other independent variables (Agresti and Finlay, 1984). The process continues until each remaining independent variable makes a significant partial contribution to explaining the variability in the dependent variable. In all the stepwise models a 5% significance level for removal of the variables was used.

For the subset of respondents who stated that they had some knowledge of IVF a multiple regression model was used to study the effects of demographic, social and economic variables on the level of knowledge of IVF, measured according to a 5 level scale. As ordinary least squares (OLS) estimates of coefficients were not efficient in the case of binary variables, a logistic model is used (Maddala, 1983). In such a model, generally called *logit*, it is assumed that errors follow a logistic cumulative distribution. For this model, goodness of fit is given by a measure called *pseudo* R^2 which is similar to the coefficient of multiple determinations (R^2) used in multiple regressions. As for R^2, pseudo R^2 ranges from 0 to 1 with higher values reflecting the better overall goodness of fit. The overall significance of the model is measured by the likelihood-ratio statistic which follows a χ^2 distribution. In addition, for all logit models we compared observed and predicted values and we calculated the Count R^2 that is the proportion of correct guesses predicted by the estimated model (Long, 1997).

In the logit models, coefficients are expressed as odds ratios, which are ratios between the odds of the two groups which characterize the sample (e.g. males and females). Odds are derived by probabilities; they are the ratio between the probability of the event and its complement to one. If the probability of an event is 0.8, its odds is 4 (0.8/0.2 = 4). Let's imagine that the probability of an event varies between males and females; it is 0.8 for females and 0.66 for males; thus the odds for females is 4 (0.8/0.2=4) and for males is 2 (0.66/0.33= 2). As the odds ratio it is simply the ratio of the two odds, in this case it would be 2 (4/2). Results of the logistic regressions are reported as odds ratios (rather than as coefficient of linear models) as they can be easily derived by the likelihood function. For all dependent variables included in the model z-scores and their p-values are reported.

5.5. Hypothetical personal use of IVF

Before asking about the WTP for using IVF we asked survey participants if they would have personally tried IVF in a hypothetical situation of infertility. More specifically, we presented the following question:

"Imagine that you have been married for a few years and you have been unsuccessfully trying to have a baby. Imagine you are told to use IVF and that it has a 30% chance to be successful. Would you personally try IVF?"

We gave respondents 5 options to this question: "do not know", "yes, definitely yes", "yes, probably", "no, probably", "no, definitely not". Similarly to the questions on attitude towards IVF, we used regression analysis to explain the variability in the willingness to try IVF in case of infertility. As the questions were framed with five possible answers ("do not know"; "yes, definitely"; "yes, probably", "no, probably" and "no, definitely") two main measurement strategies could be followed. In the first model we computed "do not know" as missing value and we dichotomised the remaining 4 answers (yes versus no). In such a model, in addition to treat the undecided as missing values, the information on difference between who is probably and who is definitely for the two options is lost. The second model tries to exploit the full information content of the 5 possible answers. Accordingly, answers are assumed to be ordered on a scale from 1 to 5 with "do not know" as the intermediate value (3) between "probably not" (2) and "probably yes" (4).

As the dependent variable in the first model is dichotomous we used a logistic regression. For the model with 5 ordered levels, measuring the intensity of the willingness to try IVF, we ran a multiple regression analysis (i.e. OLS). Similarly to the analysis reported in the previous section, for both the models we report results of three analyses: i) the "full dataset" including all possible independent variables presented above (dataset a), ii) the "full dataset" including the variables selected by a backward stepwise regression (dataset a1), iii) a model based on the "income dataset" including the variables selected by a backward stepwise regression (dataset b).

In addition to demographic (age, fertility age, gender, number of children, marital status), geographic (macro area, county town), social (education and employment) and economic variables we included the answers to the questions on knowledge about infertility and ARTs and on acquaintance with someone who had used IVF.

Coefficients estimated in the regression models can be used to predict the dependent variable (willingness to try IVF) for specific values of the dependent variables (e.g. income, education, age, etc.). Because in logistic regression the dependent variable has binary values (0 and 1), predictions are expressed as probabilities of a positive outcome (that is probabilities of willing to try IVF in case of infertility). Predictions are generated by the "income dataset" (dataset b) and the logistic regression model specified by the backward stepwise procedure.

5.6. Willingness to pay for using IVF

A specific question was framed to elicit how much respondents were willing to pay in the case they wanted to use IVF. The question was presented on computer screen to the survey panel as it approximately appears in table 5.2 (translated into English).

Table 5.2. On screen presentation of the question on willingness-to-pay for IVF

"And, would you try IVF if.......Please, for each alternative indicate yes or not"	
If it were free	□ Yes, I would try IVF □ No, I would not try IVF
If it cost £ 1 million (€ 516)	□ Yes, I would try IVF □ No, I would not try IVF
If it cost £ 5 million (€ 2,582)	□ Yes, I would try IVF □ No, I would not try IVF
If it cost £ 10 million (€ 5,165)	□ Yes, I would try IVF □ No, I would not try IVF
If it cost £ 20 million (€ 10,329)	□ Yes, I would try IVF □ No, I would not try IVF
If it cost £ 50 million (€ 25,823)	□ Yes, I would try IVF □ No, I would not try IVF

Respondents were presented with 6 values (presented in Italian Lira in the questionnaire and shown here also in Euro at the official parity of 1,936.27 It. Liras for one Euro).

Two main assumptions can be made about the rationality of respondents (potential users of IVF). The first assumption, which reflects standard economic theory, is that respondents are not expected to answer "Yes" at any given amount, if for lower amounts they have answered "No". Indeed, it looks irrational to pay a higher price for a good for which a lower price is rejected. If we make this rational assumption we can also make a consistency check to identify all the answers that violate this assumption. However, it is also plausible to assume that for some respondents higher prices may reflect higher quality. Given the peculiarity of the good (medical interventions aimed at generating babies) some respondents may distrust low prices, especially in the case they new market prices, and may be attracted by higher prices because they may signal quality. Nevertheless, even if this second assumption of rationality is accepted a consistency check can be performed. It would be inconsistent to violate twice the rationality rule described above. For example, it would be not rational for respondents to state they would not try IVF if it were free, then that they would pay It. Liras 1 million, then that they would not pay It. Liras 5 million and then, finally, that they would pay a higher amount.

To check whether the "normality" rule was violated twice we designed an algorithm that listed all records meeting the criteria shown in table 5.3. The algorithm was then translated into instructions to the STATA program to identify records that matched the criteria.

Table 5.3. Algorithm used to check the violation of rationality of answers to the questions about WTP for personal use

The question posed 6 bids (£ 0, 1, 5, 10, 20 and 50 million It Liras). If x (normal character) denotes a "yes" answer to the bid and **x** (bold character) a "no" answer, the following pattern of responses were identified as a violation of rationality according to the second assumption (see text).

0 1 **5** 10 20 50	**0** 1 **5** 10 **20 50**	**0** 1 **5** 10 **20** 50	**0** 1 **5 10** 20 50	**0** 1 **5** 10 20 **50**
0 1 **5 10** 20 **50**	**0** 1 **5 10 20** 50	**0** 1 5 **10** 20 50	**0** 1 5 **10 20** 50	**0** 1 5 **10** 20 **50**
0 1 5 10 **20** 50	0 **1** 5 **10** 20 50	0 **1** 5 **10** 20 **50**	0 **1** 5 **10 20** 50	0 **1** 5 10 **20** 50
0 1 **5** 10 **20** 50				

Although the formulation of the WTP questions may appear similar to a payment card, it is important to appreciate that respondents provided 6 different answers rather than an answer on a single value or a range of values. We have labelled this method as "Modified Payment Card" to highlight the fact that it is not a payment card, even if the data were analysed according to a method that is similar to that used for payment cards. In effect we assigned, each respondent to one value only and this value was calculated as the mean between the value for which she stated to be willing to pay and the value for which she stated not to be willing to pay.

From the answers to the WTP questions for the use of IVF it is possible to derive a demand schedule for IVF. It is worth noting that the demand curve represents the relationship between price and "number of consumers". The quantity axis is not the number of IVF used by each consumer, but rather the number of consumers using an IVF programme. Really, data collected do not define a curve but just 6 points. In order to derive a demand curve two assumptions have to be formulated. The first concerns the lines between the points. We assumed that demand is linear between each consecutive pairs of points. The second assumption, more critical, concerns the right tail of the WTP distribution (respondents with the highest WTP values). Five hundred individuals answered "yes" to the It Liras 50 million bid (€ 25,823). While there is no direct evidence that the respondents would be willing to pay even higher amounts it is very likely that some of them would. To capture this possibility we created two different scenarios. In the first one we assumed that It. Lira 50 million is the maximum amount that respondents are willing to pay; in the second scenario we assumed that the number of respondents willing to pay linearly declines at a constant rate. We also assumed that this rate is the same as the one calculated for the decline of the positive respondents between the It. Liras 20 million (€ 10,329) and It. Liras 50 million (€ 25,823). Aggregate measures of WTP can be derived from the demand schedule as total WTP is the area beneath the curve and can be calculated numerically.

The sample was built to include a minimum number of people in respect of age, gender, education, employment status, geographical area and municipality size of respondent' residence. For each stratum the number of people expected to be included in a representative sample of the Italian population was used as a basis to weight each member of the panel. Simply stated, the members of the panel who were part of underrepresented strata (e.g. elderly) were given a high weight while those over-represented were given a low weight (low education level). These weights were used to calculate mean WTP for personal use of IVF and for a national programme providing IVF with public funds.

5.7. Willingness to pay for a publicly funded programme providing IVF to infertile couples

WTP for a public programme providing IVF to infertile couples was elicited using two approaches: take-it-or-leave-it (TIOLI) and modified payment card (MPC).[12] According to the TIOLI approach, individuals are asked whether they would pay a specified amount for a given commodity/programme (Bishop & Heberlein, 1979; Ryan, 2004). The specified amount is varied across individuals so that for each respondent it is only known whether her/his WTP is above or below the bid offered. The PC approach, which was developed by Mitchel and Carson (1981, 1989), is based on the presentation of a range of bids; for each bid each respondent is asked to state if she/he would be willing to pay that amount for the commodity/programme.

[12] In effect the present study adopts a spurious version of dichotomous choice questions as respondents could graduate their answers by choosing between "probably" and "definitely" for each alternative ("yes" or "no") and by answering "do not know". Dichotomous choice questions have also been referred to in the literature as take-it-or-leave-it, closed-ended or referendum (Ryan et al., 2004). See chapter 2 for a comprehensive discussion of elicitation methods.

This method is more efficient than the DC approach as it collects more precise information about each respondent's WTP.

When we designed the survey CV research was still strongly influenced by the NOAA report which clearly stated that OE questions were to be avoided and that the TIOLI (Take-It-Or-Leave-It) format had to be preferred (Arrow et al., 1993; Smith 2006). Given that our survey had several innovative elements we decided to adhere to these recommendations and thus we opted for TIOLI without follow-up as our principle elicitation format. However, we had to limit the values used in this part of the survey to four. We were fully aware that using such a limited number of values is a major limitation and produces imprecise estimates, in the sense that the uncertainty around the estimates is very large. Nevertheless, we were forced to limit the number of bids for two main reasons. The first was mainly technical. The split of the electronic panel required the modification of software routines and this would have been too expensive for more then four bids. In addition, and more importantly, with only four sub-samples it was possible to maintain certain characteristics of each sub-sample and thus assure that we could have WTP estimates representative of the adult Italian population.

Our main objective was to estimate WTP of the Italian population so to have a welfare measure really usable to offer guidance to policy-making. Furthermore, we wanted to provide empirical evidence on methodological issues. As far as the elicitation methods are concerned, we wanted to use a second elicitation method and compare it to TIOLI. We looked at the Payment Card (PC) format, mainly because it allowed us to increase the number of values to which the sample could answer. Still now PC is a method that attracts a lot of attention, despite serious problems created by range biases and prominent numbers (Whynes et al., 2004; Whynes et al., 2005). We decided to frame our PC table as a list of values including those used for the TIOLI part and thus we added two higher and two lower values.

One special option that electronic surveys makes available is to block respondents if they do not answer to questions. Basically, the respondent to the survey can be prevented to pass to the successive questions if the answer box is not filled-in. Clearly, this is an important feature because it reduces item non-response rate and thus reduces the risk of having a biased sample of respondents. On the contrary, however, forcing answers may increase the number of protesters thus introducing more uncertainty in the estimates and, possibly, other biases. Given the objective of the thesis, that was to provide the best possible welfare estimate of a national programme publicly funding IVF, we opted for forcing respondents to answer to all questions and then checked for protest answers.

The internet-based method provided us with an additional feature that is rather unexplored: to randomly generate the sequence of the eight values. In other words, the software managing the delivery of the questionnaire could generate, for each respondent, a particular order of the values presented. This randomization allowed us to limit range bias because respondents do not see the minimum value at first and they may not be aware that they had selected the maximum value when checking it. In addition, randomization should eliminate the risk that the sequence of questions may have any effect on answers.

As a consequence of this feature we also opted for an additional change of the standard way in which payment card questions are generally formulated. Instead of presenting a list of ranges we opted for asking yes/not questions for specified values. Once the respondents are provided a casual order of questions on specific values we thought to be more appropriate to use yes/not questions to specific values rather than selecting a specific range of values. Our solution is simpler and less cognitively demanding. Since this method is different from the standard PC method, we called it the modified payment card (MPC).

From the point of the analysis of data, the choice of asking several yes/not questions has important implications and casts doubts about the type of elicitation format we used. First of all we have 8 observations for each respondents and not only one has it is in the standard PC format. This allows us to use the data at least in two ways. With the first method we can calculate the number of "yes" respondents for each value and estimate a survival function to which derive, parametrically or non parametrically, median and mean WTP. The second approach is to derive from each respondent the highest WTP and use these values to parametrically or non parametrically estimate mean and median WTP. Provided that we use the same assumptions about the underlying distribution of data and the algorithms used, we expect to obtain similar estimates and to attribute difference to errors and protest answers.

Individuals were first presented the TIOLI question. After the question on WTP for using IVF it was given a short illustration of a hypothetical referendum on this issue. The referendum was illustrated in the following way:

> *"Imagine that there is a referendum in order to decide whether to fund IVF with public money. Each year public funding would be available to about 30,000 couples who are advised to use IVF. Out of these couples about 10,000 would have a baby. Without public funding, couples requiring IVF should spend between It. Liras 5 and It. Liras 15 million for the procedure"*[13].

Then, individuals were asked to state their vote. The total sample was randomly split in four sub-samples and for each of them a different annual payment (termed "tax") was proposed. The question was framed as follows:

> *"Would you vote in favour of funding IVF with public money if you were asked an annual payment, that is a tax, of Italian Liras X?""*

The following four values were proposed: Italian Liras 5,000 (€ 2.6), It. Liras 10,000 (€ 5.2), It. Liras 20,000 (€ 10.3) and It. Liras 50,000 (€ 25.8). Respondents could choose between five options: yes, definitely; yes, probably; do not know; no, probably not; and no, definitively not.

The MPC questions were the same for all respondents and were presented after the TIOLI question. Respondents were given 8 values from It. Liras 2,000 (€ 1) to It. Liras 200,000 (€ 103.3) (table 5.4).

[13] Italian Liras 5 and 15 million correspond to € 2,582 and € 7,745, respectively.

Table 5.4. Modified payment card willingness-to-pay questions.

"How much would you be willing to pay annually? For each alternative, please indicate would you try IVF if.......Please, for each alternative indicate yes or no."	
It. Liras 2,000 (€ 1.0)	□ Yes, I would pay □ No, I would not pay
It. Liras 5,000 (€ 2.6)	□ Yes, I would pay □ No, I would not pay
It. Liras 10,000 (€ 2.6)	□ Yes, I would pay □ No, I would not pay
It. Liras 20,000 (€ 10.3)	□ Yes, I would pay □ No, I would not pay
It. Liras 50,000 (€ 25.8)	□ Yes, I would pay □ No, I would not pay
It. Liras 80,000 (€ 41.3)	□ Yes, I would pay □ No, I would not pay
It. Liras 100,000 (€ 51.6)	□ Yes, I would pay □ No, I would not pay
It. Liras 200,000 (€ 103.3)	□ Yes, I would pay □ No, I would not pay

Values were presented randomly to each respondent. The values were purposely chosen to include the same amounts offered in the TIOLI questions and to cover a larger range. For each value respondents were required to choose between *yes* and *no*, without the "*do not know*" option and the possibility to graduate the answers by the use of *probably* and *definitely*.

5.8. Demographic and socio-economic determinants of WTP for a publicly funded programme providing IVF to infertile couples

We used various regression models to identify significant predictors of WTP for the publicly funded programme providing IVF to infertile couples. For the TIOLI format we used two types of models. In the first type we computed "*do not know*" as missing values and dichotomised the remaining 4 answers (yes versus no). In such a model, in addition to treating the undecided as missing values, the information on the difference between who is "probably" and who is "definitely" willing to pay the proposed amount is lost. This coding strategy allowed us to analyse the data by logistic regression (Long, 1997). The second model tried to exploit the full information content of the 5 possible answers. Accordingly, answers are assumed to be ordered on a scale from 1 to 5 with "*do not know*" as the intermediate value (3) between "*probably not*" (2), "*probably yes*" (4) and the two definite answers as extremes (1 and 5). To analyse this data we used ordered logit models (Long, 1997). In addition to demographic and socio-economic variables in these models the value of bids is added as one the independent variables.

Similarly to the regressions presented previously, three specifications are presented: i) based on the "full dataset" (dataset a) including all possible demographic and socio-economic variables, ii) based on the same dataset where variables are selected by a backward stepwise procedure (dataset a1), iii) and based on the "income dataset" where variables are selected by a backward stepwise procedure (dataset b).

We also ran regression analyses for the data collected through the MPC questions. Each respondent had to circle whether he/she *was* or *was not* willing to pay 8 different amounts. Answers to these questions were coded as maximum WTP: for each respondent it was coded the highest amount for which he/she was providing a "yes" answer.

In the presence of censoring OLS estimates are inconsistent (Maddala, 1983). The appropriate approach in such circumstances may to estimate a *tobit model*, something referred to as the *censored regression model*. The tobit model uses all the information, including information about the censoring, and provides consistent estimates of the parameters through a maximum likelihood (ML) procedure (Long, 1997). ML estimation for the *two-limit tobit model* involves dividing the observations into three sets. The first set contains uncensored observations, which ML treats in the same was as the Linear Regression Model. The other two sets contain censored observations. For these observations, we do not know the specific value of the dependent variable because of censoring, but can proceed by computing the probability of being censored and using this quantity in the likelihood equation.

As for the data generated by the TIOLI question we ran one regression with the entire dataset including all the possible demographic and socio-economic variables and two backward stepwise regression models, one using the entire dataset and the other using the sub dataset containing valid observations of the self reported monthly family income after taxes.

5.9. WTP estimates: the referendum format

As in the tradional referendum (TIOLI) question respondents could graduate their answers choosing between *definitely* and *surely* and could answer *do not know*, various approaches can be followed to transform data in a binary format and to estimate mean and median WTP. Here we present four possible approaches: i) to combine *definitely* and *probably* answers and treat "do not know" as missing values; ii) to combine *definitely* and *probably* answers and treat "do not know" as negative answers; iii) to consider only "yes, definitely" as "yes", both *definitely* and *probably* negative answers as "no" and "yes, probably" and "do not know" as missing value; iv) to consider only "yes, definitely" as "yes" and all the other answers, including "do not know", as negative answers. Approach i) is the most favourable to the programme; it is expected to generate the highest estimates of WTP. On the contrary, approach iv) is the most conservative as it assumes that all answers but "yes, definitely" reveal negative WTP for the IVF programme. Approaches ii) and iii) are expected to produce less extreme estimates. Table 5.5 summarises the assumptions of the four scenarios.

Table 5.5. Approaches to estimate willingness to pay for the IVF programme elicited through the take-it-or-leave-it format

Assumptions	Original answers stated by respondents					Formulas to calculate WTP
	Yes, definitely (a)	Yes, probably (b)	No, definitely (c)	No, probably (d)	Do not know (e)	
1) Least conservative	Yes	Yes	No	No	Missing	(a+b)/(c+d)
2) intermediate	Yes	Yes	No	No	No	(a+b)/(c+d+e)
3) intermediate	Yes	Missing	No	No	Missing	(a)/(c+d)
4) Most conservative	Yes	No	No	No	No	(a)/(b+c+d+e)

We assumed that the probability that an individual will say "yes" to the WTP question follows a logistic distribution (Ryan, 1998 and 2004):

$$P(yes) = (1 + e^{-a+b*bid})^{-1}$$

where a and b are the coefficients of the estimated regression logit equation when only the bid amount (bid) is included and P(yes) is the probability of accepting the bid. Following Hanemman (1984) and Ryan (1998 and 2004) mean WTP can be estimated as the area under this probability function. This area shows the fraction of the sample who would consume the good at each price (bid) level. Therefore mean WTP can be estimated by integrating the probability function.

$$E(WTP) = \int_{L}^{U} (1 + e^{-a+b*bid})^{-1} db - \int_{L}^{0} 1 - (1 + e^{-a+b*bid})^{-1} db$$

Where U is the upper limit of the integration and L is the lower limit of the integration.

Mean WTP is thus critically dependent on making assumptions about the upper and the lower limits of the integral. Assuming that WTP is monotonic these limits should correspond to the values for which probability of "yes" is zero (the lower limit) and probability of yes is one (the upper limit). Both of them are not observed in the present study. As far as the lower limit is concerned, 31.5% of the total sample was not willing to pay Italian Liras 5,000 (€ 2.60), the lowest bid offered (see next chapter). A lower limit has thus to be assumed. About 41% of respondents were definitely willing to pay Italian Liras 50,000 (€ 25.80), the highest bid offered. Consequently, imposing this value as the upper limit is conservative and assuming higher values may be appropriate.

We made two different assumptions about the upper limit: Italian £ 50,000 (€ 25.80), that is the highest value for which some "yes" answers are observed and Italian Liras 200,000 (€ 103.30) that is the highest value offered in the MPC question.[14] For contingent valuation studies in the healthcare sector zero is generally used as the lower limit because it can be assumed that individuals will not use the healthcare intervention if they receive disutility from it. This assumption is plausible when individual's utility does not include caring externalities, that is the utility of one person is not affected by the consumption of other persons. If the utility of an individual is affected by others' consumption of the good the sign of the contribution cannot be assumed a priori. This is because the person whose utility is affected by others' consumption of the good does not control whether the good is used or not (as it is the case if he/she is the user of the good). While this argument is often not relevant in

[14] It should be noted that the modified payment card questions were presented immediately after the TIOLI one that generated the data analysed in this section.

health care as individuals are not expected to get disutility by the use of health services by others, it may be relevant in the case of Assisted Reproductive Technologies. Since these technologies manipulate embryos and interfere with natural reproduction some people are ethically against them and may suffer utility losses from the use of these technologies by other members of the community. Following this argument we decided not to completely rule out the possibility of a negative lower limit of the integration. We performed the computation of mean WTP for IVF according to two lower limit assumptions: 0, the standard assumption in the health economics literature and a negative value of It. Liras 50,000 (€ 25.8).

Environmental economics literature suggests that there may be individuals that have zero WTP for a public good as they derive no value from consumption of the good. Indeed, some individuals may suffer from the provision of specific public goods. We designed the survey in such a way that respondents could clearly state that they were not in favour of any public funding of IVF. Basically, they could attribute no value to a public programme providing IVF and thus it can be assumed that they exactly have zero WTP. The presence of these individuals makes critical the use of a logistic function to describe respondents' WTP because, by definition a logistic function rules out a nonzero probability of zero WTP (Kristrom, 1997). The function assumes that all respondents are in-the-market for a public programme providing IVF, but this is not true according to our data and intuition.

To take into account zero WTP we used a basic spike model suggested by Kristrom (1997) and also used by Yoo and Kwak (2002).

We assumed that the distribution function of WTP is given by:

$$F_{wtp}(A) = \begin{cases} 0 & \text{if } A<0 \\ p & \text{if } A=0 \\ G_{wtp}(A) & \text{if } A>0 \end{cases}$$

Where $G_{wtp}(A)$ is a function such as that $G_{wtp}(0) = p$ and $\lim_{\rightarrow \infty} G_{wtp}(A) = 1$.

Intuitively, WTP has zero probability for negative values, probability p for WTP = 0 and probability described by $G_{wtp}(A)$ if WTP is positive.

This model can be estimated through parametrically and non-parametrically methods. A non-parametric model was used following the formula suggested by Kristrom (1997). In addition we estimated a parametric model under the assumption that $G_{wtp}(A)$ follows a logistic probability distribution. These spike models were compared to "traditional" models to check the relevance of different assumptions concerning the truncation of the distribution at zero. In the parametric case, under the assumption of a logistic distribution, the spike of the distribution can be found at $1/(1+\exp(\alpha)$, where α is the constant estimated by the maximization of the logit function and the mean is $\log((1+\exp(\alpha))/\beta$, where β is the coefficient of the bid term.

5.10. WTP estimates: the modified payment card format

After the referendum question, respondents were asked to state if they were willing to pay for the IVF programme 8 different amounts. For each amount, presented in random order, respondents were requested to sign "yes" or "no". To calculate mean WTP from these data we adopted a standard procedure which assumes true WTP as the mid-point between the amount at which the respondent said "yes" and the amount at which he/she said "no" (Cameron, 1987; Ryan, 2004). This approach allows WTP to be estimated without regression techniques. Therefore, it does not require any hypothesis about the distribution of the underlying probability function. However, it requires making assumptions about minimum and maximum WTP. Minimum WTP is assumed to be 0; therefore respondents who were not willing to pay It. Liras 2,000, the minimum vale offered, were attributed 0 WTP. For the people willing to pay It. Liras 200,000 (€ 103.3) it was assumed that this was their maximum WTP. Both these assumptions are conservative, as they underestimate true WTP. However, 0 minimum WTP rules out the possibility of negative WTP.

The method to calculate mean WTP can be also illustrated graphically and it is substantially identical to that used for estimating mean WTP for personal use of IVF. Data can be used to present a pseudo-demand schedule for the programme. Amounts offered to respondents are presented in the horizontal axis and the number of respondents willing to pay those amounts are presented in the vertical axis. For the two extreme values it was assumed that minimum and maximum WTP was 0 and

It. Liras 200,000, respectively. It was also assumed that demand was linear between each consecutive pairs of points. The area below the curve represents the total WTP of the entire sample and can be calculated numerically.

5.11. Validity analysis and comparison of the two elicitation methods.

The referendum question was posed before the MPC questions. It is thus possible to check whether the answer to the first question had any impact on the answers which followed. Since the four groups are not significantly different in their demographic and socio-economic characteristics, any difference in mean WTP or other estimates derived from results of the MPC should be attributed to the answer to the TIOLI question. We compared mean WTP elicited from the four groups. We made two comparisons: with all data included to calculate mean WTP and excluding "do not know" answers. The two comparisons reflect two different sources of bias. In the former comparison, the source would be the value offered; respondents are anchored to the value offered in the TIOLI question. In the latter comparison, the source of bias would be the decision concerning accepting or rejecting the value offered. In this validation check, "do not know" answers would not be a source of bias and thus should be excluded by the analysis.

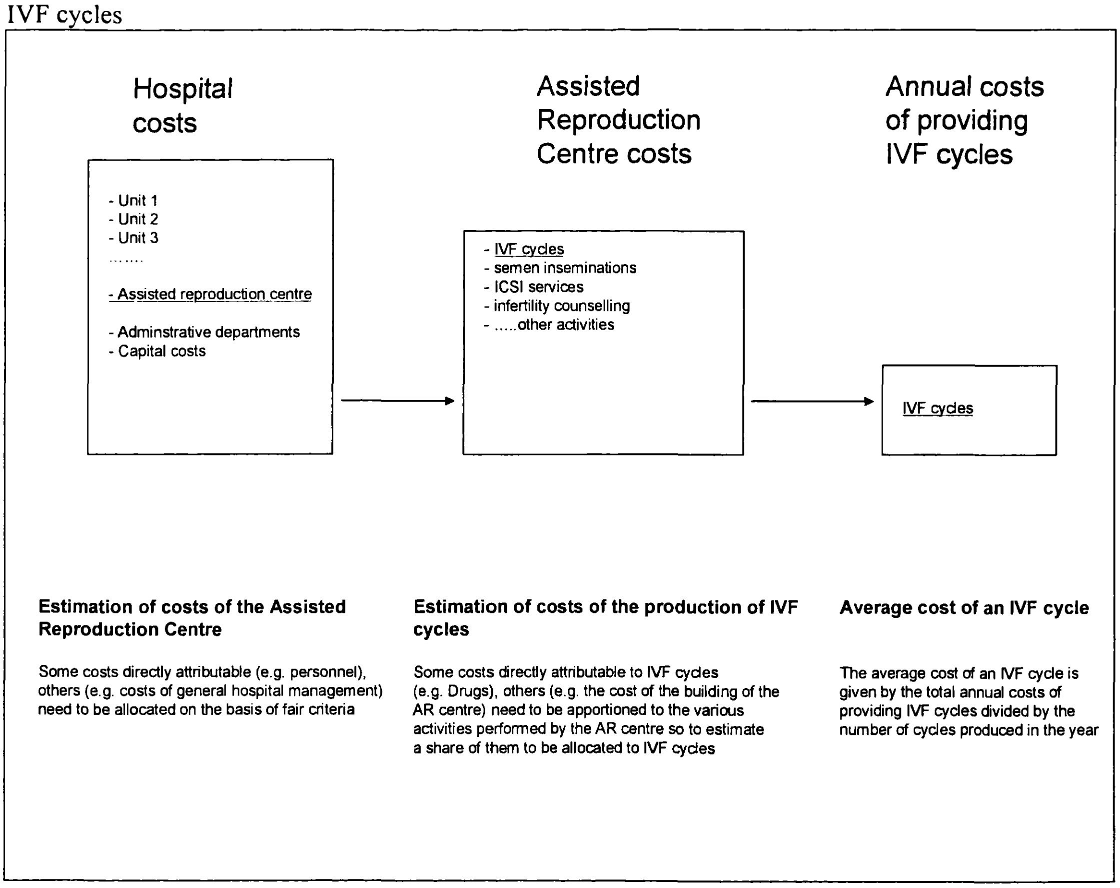
We first compared mean WTP across the two elicitation methods. Then, we transformed TIOLI data in MPC data and compared them to the original MPC data. As mean WTP generated from TIOLI data requires various assumptions about integration limits and the density function of the probability distribution, differences of estimates between the two datasets may derive from these assumptions. To neutralise the implications of these assumptions, we generated virtual TIOLI data from the data collected through the MPC and, in order to calculate mean WTP, we made the same assumptions used for the original TIOLI data. In order to do this we created a new binary variable (yes and no). For each respondent it was first detected to which of the four groups used for the TIOLI question he/she was attributed. Then, it was compared the maximum WTP value elicited from MPC data with the bid attributed to that group. If the former was equal or larger than the latter the respondent was attributed "yes", otherwise it was attributed "no". For example, if the respondent was in the It. Liras "20,000" group he/she was coded as "yes" if in the MPC question he/she reported to be willing to pay It. Liras 20,000 or above and he/she coded as "no" if his/her stated WTP in the TIOLI was lower than this amount.

5.12. Methods used to estimate costs of production and cost-effectiveness of In-Vitro-Fertilization

Cost data were mainly derived from management accounting documents and were complemented by interviews with clinicians and administrative staff employed by two Assisted Reproduction (AR) centres based in two major hospitals, one private and one run by the Italian NHS. Both of them are located in Lombardy, the largest and most populated region in Italy. Data was collected in 1999 and refer to cost and activity data in fiscal year 1998. In that year the private and the NHS run centres performed 2,257 and 1,116 IVF cycles, respectively.

On the basis of arguments presented in the previous section, we used a full cost accounting methodology (Horgren et al., 1996). First, with the help of the administrative offices we estimated total annual costs of the organisational units managing the assisted reporduction (AR) centres (figure 5.1). Then, we estimated the part of these costs that are absorbed to provide IVF cycles as the AR centres produce a variety of services in addition to IVF. Finally, we divided the annual costs of providing IVF services by the number of cycles delivered.

Figure 5.1. Rappresentation of the methodology used to estimate the average full cost of providing an IVF cycles



In both organisations management accounting data attributed only direct costs to the AR centres. Indirect costs remain unallocated by the management accounting systems in the sense that the organizations do not calculate any apportioning of overheads to organizational units. Direct costs consider the personnel working for the unit, drugs, medical devices, diagnostic services provided by other departments, use of the operating room, ward care (day hospital) and utilities (table 5.6).

Table 5.6. Cost categories and method of estimation

Cost categories	Methods of analysis
Personnel	Direct estimate based on administrative data (cost of labour) and allocation of professional time to the IVF centre
Drugs	Direct allocation based on administrative data
Disposables and other material	Direct allocation based on administrative data
Equipments (amortization)	Direct allocation based on administrative data (private centre); estimation based on costs of acquisition and straight depreciation (public centre);
Diagnostic services	Direct allocation based on administrative data (transfer prices)
Operating room	Direct allocation based on a hourly cost and the total amount of time of operating theatres used by the IVF
Hospital care	Direct allocation based on administrative data (average cost of day hospital care net of operating theatre, diagnostic and drug costs)
Utilities	Percentage of total hospital costs
Other	Percentage of total hospital costs
Overheads	Percentage of total hospital costs
Building (amortization)	Direct allocation based on administrative data (private centre); estimated on the basis of market rents

For the private infertility centres accounting data also reported the amortization value of health technologies used by the AR centre and of the building where the centre is located. For the public centre the data was not available and were estimated upon qualitative and quantitative information collected during the interviews. The annual cost of the building used by the centre was estimated on the basis of market rents of buildings for commercial use. The value of health technologies available at the AR centre was calculated through a straight depreciation procedure of the acquisition costs assuming an average useful life of the technology of 5 years. For both centres a fair share of general costs to be attributed to the AR centre was calculated on the basis of the ratio between the labour costs of the AR centre and the total labour costs of the hospital.

This is a rather crude method to allocate general overheads but is commonly used in many organizations. This method implicitly assumes that the ratio between total direct and indirect costs of the AR centre can be approximated by the ratio between the value of the labour directly attributable to the AR centre and the total value of the labour costs of the hospital. More formally, let CL and IC the costs of labour and indirect costs (the total amount of overheads), respectively. If the AR suffix refers to the Assisted Reproduction centre and HOS identifies the entire hospital:

$\theta = CL_{AR} / CL_{HOS}$

where θ is known because both costs are given by the hospitals. The assumption is that θ is also the share of the total general costs that has to be attributed to the AR centre. Since θ and the amount of total hospital costs are known, the amount of indirect costs to be attributed to the AR centre is:

$IC_{AR} = IC_{TOT} * \theta$

The second step of the procedure is to estimate the part of resources attributed by the AR centre that are used to provide IVF service. This is necessary because AR centre produces a variety of services in addition of IVF cycles.

In both centres the costs of diagnostic services, drugs, the use of the operating theatre and of ward care (the cost of the day hospital for the intervention) could be directly attributed to IVF cycles. For these items was thus possible to directly calculate the total amount of costs attributable to IVF cycles. For the other cost components it was not possible to make a direct allocation. For each category of these costs, the percentage of the total consumption of resources that could be attributed to IVF cycles was asked to our informants. Basically, for each category of costs we estimated the part of them attributable to the production of IVF services. As for overheads, capital costs and utilities informants could not provide any reasonable answer, we apportioned these costs to IVF on the basis of the average percentage estimated for the other categories of indirect costs. For each of the two centres, all direct and indirect costs of providing IVF services were added. Then the cost per IVF cycle was calculated by dividing this total by the number of cycles provided

These cost estimates do not include the costs of treatment with gonadotrophin-releasing hormone (GnRH) and with Follicle-stimulating hormone (FSH). The costs of these treatments for each IVF cycle were estimated on the basis of the protocols used by the two AR centres.

These calculations were used to estimate the average cost per IVF cycle. However, complications associated with the procedure may be a cause of additional costs. Infertility treatments may cause severe ovarian hyperstimulation, other major side effects (Neumann, et al., 1994) and multiple births (Schenker, 1994; Human Fertilisation and Embriology Authority, 2002). Neumann et al. (1994) estimated that the additional cost of complications (side effects of treatment, higher rates of difficult pregnancies, complications attributable to multiple births) is US $ 1,717 per IVF cycle. This value is approximately 20% of the cost of one cycle of IVF (US $ 8,000) as estimated by the same authors. This percentage was applied to the cost per IVF cycle estimated in our study in order to get an estimate of the expected costs of complications for each cycle of treatment.

Given the relevant uncertainty surrounding the data, we performed a sensitivity analysis of cost data by using wide range of costs per IVF taken from an Italian study that provided estimates for each of the major Italian regions. The minimum and maximum regional value of cost per IVF cycle was used to perform a sensitivity analysis of the incremental cost-effectiveness ratio (Mantovani et al., 1999).

The attribution of costs of complications to each cycle of treatment, as well as the presentation of basic information to the respondents of the IVF survey, require estimates of the success rate of IVF. These estimates can be referred to clinical pregnancy rates or live birth rates. The cost-benefit analysis presented in this study focuses on clinical outcomes and is based on a contingent valuation survey of the WTP for procreating babies through IVF. Therefore the relevant measure of success is live birth. As Italian data on IVF cycles and outcome was not available at the time of the analysis, effectiveness of IVF was derived by the World Register for year 1993 (De Mouzon and Lancaster, 1995). The register reports the pregnancy rate and the delivery rate per egg retrieval. As only a fraction of initiated IVF cycles results on successful egg retrieval (Newman et al, 1994), delivery rates reported in the World Register were adjusted in order to obtain the probability of live birth per initiated IVF cycle.

Cost and effectiveness data was used to formulate the scenario for the contingent evaluation survey and to calculate cost per live birth. Following Neumann et al. (1994), it was assumed that the probability of delivery declines one percentage point per cycle. This is because the women with the higher potential for fertility are more likely to become pregnant early; consequently those women who keep trying IVF have a lower probability of becoming pregnant. Consistently with the Italian legislation about fertility treatments, a maximum of three IVF cycles were taken into consideration and the possibility of freezing eggs was excluded.

The pregnancy rate and delivery rate varies across clients of Assisted Reproduction centres. It is thus important to go beyond the analysis to the "average" patient and to estimate cost per delivery according to different scenarios. US 1993 data on 31,418 cycles of IVF were used to estimate delivery rates for four categories of patients: i) under 40 with male infertility factor, ii) under 40 without male infertility factor, iii) over 40 with male infertility factor, iv) over 40 without male infertility factor. Although indications of IVF are not analysed in the cost-benefit analysis is worth remembering that the rate of success and thus the economic value of the intervention greatly vary according to patients' age and the diagnosis of infertility.

5.13. Cost-benefit analysis

On the basis of estimated costs and benefits we suggest a few analyses that can offer guidance to policy making. First, we simply subtract mean WTP for personal use of IVF to mean cost, to obtain the net benefits of personal use of IVF. We calculated mean WTP per respondent and per user. However, since IVF would generate costs only for those willing to use it, the correct cost-benefit analysis should exclude non users.

The main goal of this study is to assess the provision of IVF from a societal perspective. Therefore, the main cost-benefit analysis is based on the benefits perceived by a sample of the national population for a programme providing free access to IVF to infertile couples desiring to have a baby. On the basis of the scenario presented in the survey and the results of the CV answers, we estimated aggregate benefits of a national programme providing publicly funded IVF to 10,000 infertile couples to obtain 3,000 babies. In order to obtain these estimates we extrapolated aggregate WTP of the sample to the entire Italian adult population. Then, the national aggregate WTP was compared to total costs of providing IVF to 10,000 couples.

As part of the WTP data was collected according to a referendum format, we also present results according to majority decision rules. Give the relevant fraction of the population that would never try IVF and the high WTP of a minority of populations, mean and media WTP may diverge considerably and an efficient use of economic resources may be in contrast with the will of the majority of “voters”.

Chapter 6

Knowledge, attitude and willingness to pay for In-Vitro-Fertilisation

6.1. Introduction

This chapter analyses data on knowledge of infertility and IVF, acquaintance with infertile couples, attitude towards IVF and willingness to pay for the use of IVF and for a publicly funded national programme providing IVF. The rationale of methodological choices and details about data and methods are presented in the previous chapter. Here we report the characteristics of the sample and describe the results of all the main answers of the questionnaires. Some of the results of the regression analyses are also presented and discussed. The aim of this part of the theisis is to understand how respondents' characteristics influence WTP for IVF.

6.2. Respondents' socio-economic characteristics

The return rate of the survey was 89.1% (5,739 out of 6,435). As all respondents completed the questionnaire in all parts, the return rate coincides with the response rate. High response rates are normal in this type of surveys as panel members agree in advance to be regularly surveyed and to respond to each questionnaire. It can be assumed that most of the people who failed to respond were not in the position to do so when the survey was administered (e.g. because they did not have access to the computer during the week).

Table 6.1 presents socio-demographic characteristics of respondents broken down by gender. The mean age of respondents is 40.1 years; male respondents are significantly older than female respondents (41 years vs. 39.2; t=5.06, p-value<0.0001). The number of females is slightly higher than the number of men (51.3% vs. 48.7%). About 70% of the respondents were either married or cohabiting with a partner. Only 8.5% of the sample stated that they had a university degree (at least 4 year of university studies after high school)[15] and about 10% stated having less than 8 years of education, the minimum number of years of schooling required by the Italian legislation since 1948. As expected, women were less educated than men; fewer women than men had completed high school (at least 13 years of schooling) and obtained a university degree.

Table 6.1. Socio-demographic characteristics of survey respondents (n=5,739).

	Male		Female		Total	
	#	%	#	%	#	%
Gender						
Male					2,792	48.65
Female					2,947	51.35
Age						
18-24	392	14.04	469	15.91	861	15.00
25-34	524	18.77	651	22.09	1,175	20.47
35-44	749	26.83	825	27.99	1,574	27.43
45-54	669	23.96	637	21.62	1,306	22.76
55-64	347	12.43	267	9.06	614	10.70
>64	111	3.98	98	3.33	209	3.64
Marital Status						
Married (or living with the partner)	1,998	71.56	1,999	67.83	3,997	69.65
Unmarried	794	28.44	948	32.17	1,742	30.35
Education						
Less than 8 years	249	8.92	360	12.22	609	10.61
Between 8 and 13 years	1,004	35.96	1,064	36.10	2,068	36.03
High School Diploma	1,289	46.17	1,288	43.71	2,577	44.90
University Degree	250	8.95	235	7.97	485	8.45
Employment						
Employed	1,974	70.70	1,204	40.86	3,178	55.38
Unemployed	190	6.81	250	8.48	440	7.67
Housekeeper	1	0.04	899	30.51	900	15.68
Retired	308	11.03	182	6.18	490	8.54
Student	319	11.43	412	13.98	731	12.74
Geographical Area						
North-west	677	24.25	686	23.28	1,363	23,75
North-east	384	13.75	423	14.35	807	14.06
Centre	490	17.55	538	18.26	1,028	17.91
South	838	30.01	879	29.83	1,717	29.92
Isles	403	14.43	421	14.29	824	14.36
Size of municipality of residence						
< 5,000 inhabitants	403	14.43	440	14.93	843	14.69
5,001-20,000 inhabitants	720	25.79	738	25.04	1,458	25.41
20,001-50,000 inhabitants	453	16.22	484	16.42	937	16.33
50,001-100,000 inhabitants	415	14.86	426	14.46	841	14.65
> 100,000 inhabitants	801	28.69	859	29.15	1,660	28.92

[15] As the survey was administered in 2000 no respondents could have got the three year Laurea degree introduced by the Italian legislation in accordance with the Bologna declaration.

About 55% of the surveyed individuals were employed, while the retired and unemployed accounted for 8.5% and 7.7% of the total sample, respectively. There was a 30% absolute difference in the employment rate between males and females (70.7% vs. 40.9%). Fewer women than men reported to be retired while more men than women reported to be unemployed and students. Almost 29% of the sample lived in a municipality with more than 100,000 inhabitants, while almost half of the respondents were resident in small-medium size municipalities (from 5,000 to 100,000 inhabitants). There were not significant differences between males and females in this respect.

With regard to socio-economic status, seventy percent of respondents reported to be in the three central classes (low-middle, middle, middle-upper) of the five classes according to which respondents were asked to allocate themselves. As to the request to reveal their family monthly income after taxes choosing between 13 brackets, approximately sixteen percent of respondents stated that they preferred not to answer this question (an option openly given in the questionnaire). Slightly more than 20% of the total sample responding to this question reported a monthly family income (after taxes) of It. Liras 2 million or less (about €1,033)[16]. About 8.4% of the total sample reported an income above It. Liras 5 million (€ 2,582). Females and males provided very similar answers. The two questions on the socio-economic status are associated (p-value < 0.0001), even if the positive association was not strong (Tau-b = 0.0421).

[16] In 2002 Liras were converted into Euros at the exchange rate of 1936.27 Liras for one Euro.

Table 6.2. Socio-economic characteristics of survey respondents (n=5,739).

	Male			Female			Total		
	#	%	Cum %	#	%	Cum %	#	%	Cum %
Self-reported socio-economic status									
Low	601	21.53	21.53	634	21.51	25.24	1,235	21.52	21.52
Low-middle	459	16.44	37.97	483	16.39	37.90	942	16.41	37.93
Middle	840	30.09	68.05	893	30.30	68.20	1,733	30.20	68.13
Middle-upper	682	24.43	92.48	687	23.31	91.52	1,369	23.85	91.98
Upper	210	7.52	100.00	250	8.48	100.00	460	8.02	100.00
Self-reported household monthly income (after taxes)									
Up to It £ 800.000 (€ 413)	21	0.75	0.75	26	0.88	0.88	47	0.82	0.82
It £ 800.001-1.000.000 (€ 413-516)	30	1.07	1.83	34	1.15	2.04	64	1.12	1.93
It £ 1.000.001-1.500.000 (€ 516-775)	133	4.76	6.59	139	4.72	6.75	272	4.74	6.67
It £ 1.500.001-2.000.000 (€ 775-1.033)	385	13.79	20.38	418	14.18	20.94	803	13.99	20.67
It £ 2.000.001-2.500.000 (€ 1.033-1.291)	436	15.62	36.00	468	15.88	36.82	904	15.75	36.42
It £ 2.500.001-3.000.000 (€ 1.291-1.549)	316	11.32	47.31	336	11.40	48.22	652	11.36	47.78
It £ 3.000.001-3.500.000 (€ 1.549-1.808)	261	9.35	56.66	275	9.33	57.55	536	9.34	57.12
It £ 3.500.001-4.000.000 (€ 1.808-2.066)	274	9.81	66.48	271	9.20	66.75	545	9.50	66.61
It £ 4.000.001-5.000.000 (€ 2.066-2.582)	246	8.81	75.29	256	8.69	75.43	502	8.75	75.36
It £ 5.000.001-6.000.000 (€ 2.582-3.099)	121	4.33	79.62	128	4.34	79.78	249	4.34	79.70
It £ 6.000.001-7.000.000 (€ 3.099-3.615)	41	1.47	81.09	45	1.53	81.30	86	1.50	81.20
It £ 7.000.001-8.000.000 (€ 3.615-4.132)	27	0.97	82.06	21	0.71	82.02	48	0.84	82.04
More than It £ 8.000.000 (€ 4.132)	41	1.47	83.52	57	1.93	83.95	98	1.71	83.74
" I prefer not to answer"	460	16.48	100.00	473	16.05	100.00	933	16.26	100.00

6.3. Respondents' knowledge of infertility and Assisted Reproductive Techniques (ARTs)

Almost 43% of respondents stated that they knew of couples who have infertility problems (table 6.3). Female respondents were more likely to know such couples than male respondents (44.7% versus 40.4%, χ^2 = 10.6, p-value= 0.001).

Only 779 respondents (13.6%) reported that they had never heard of In Vitro Fertilisation before. TV programmes and newspapers were the main source of information for the vast majority of the sample (72.6%). Only 82 respondents (1.4%) stated that they had been informed by doctors; these individuals were likely to have a direct personal interest in ARTs. Of the 4,960 respondents who stated having previous knowledge of IVF, about 55% reported having limited knowledge. The questionnaire did not include a specific question on personal use of IVF or other ARTs. However, it asked whether the respondent personally met someone who used them. Fifteen percent of the sample provided an affirmative answer to this question. For all questions there are significant differences between males and females as the latter were more likely to know infertile couples, to know about IVF and to have met someone who used IVF or other ARTs. Overall women looked closer than men at the issues discussed in this study.

Table 6.3. Respondents' knowledge about infertility and Assisted Reproductive Techniques

	Males	Female	χ^2 (p-value)	Total
Are you acquainted with any couples who desire to have babies and cannot have them?				
Yes	1,129 (40.44%)	1,317 (44.69%)	10.6019 (p=0.01)	2,446 (42.62%)
No	1,663 (59.56%)	1,630 (55.31%)		3,293 (57,38%)
Did you know In-Vitro-Fertilisation before?				
No, It is the first time I hear of it	406 (14.54%)	373 (12.66%)	24.8121 (p<0.001)	779 (13.57%)
Yes, I have heard of it from the Tv and/or newspapers	2,058 (73.71%)	2,110 (71.60%)		4,168 (72.63%)
Yes, I have heard of it from relatives/friends	182 (6.52%)	289 (9.81%)		471 (8.21%)
Yes, I have got informed through specialised readings	109 (3.90%)	130 (4.41%)		239 (4.16%)
Yes, I have got informed from the doctor	37 (1.33%)	45 (1.53%)		82 (1.43%)
How would you rate your knowledge about In-Vitro-Fertilisation?				
Very limited	464 (19.45%)	408 (15.85%)	31.8930 (p<0.001)	872 (17.58%)
Relatively limited	947 (39.69%)	916 (35.59%)		1,863 (37.56%)
Sufficient	821 (34.41%)	1,064 (41.34%)		1,885 (38.00%)
Relatively wide	121 (5.07%)	152 (5..91%)		273 (5.50%)
Very wide	33 (1.38%)	34 (1.32%)		67 (1.35%)
Have you ever met someone who used IVF or other ARTs?				
Yes	354 (12.68%)	537 (18.22%)	33.5858 (p<0.001)	891 (15.5%)
No	2,438 (87.32%)	2,410 (81.78%)		4,848 (84.5%)

Socio-economic characteristics are expected to explain respondents' knowledge of infertility and ARTs. Tables 6.4-6.7 present the results of a few statistical models aimed at identifying the impact of demographic, social and economic characteristics of respondents on their acquaintance with infertility couples (table 6.4), knowledge of IVF (table 6.5 and table 6.6) and acquaintance with someone who used IVF (table 6.7).

For all cardinal and ordinal variables we visually tested normality and we checked transformations. Income, socioeconomic status and years of education have such shapes that assuming a Normal distribution appears plausible. Also, when we tested non linear forms for these independent variables we have never improved the goodness of fit of the model (R^2 and likelihood ratios). Consequently, we keep these variable untrasformed in all models. Instead, for age of the respondent we consistently get better performance with a quadratic transformation, that mean that young respondents and old respondents tended to have similar knowledge and attitude, significantly different from those of the central cohorts. Consequently, in all models we used the following form for the age variable: $\beta_1(\text{age}) + \beta_2{}^2(\text{age})$.

Several variables are associated with respondents' acquaintance with couples with infertility problems (table 6.4). The most significant variable is marital status: respondents who were married (or live with a partner) have better chances of acquainting a couple with infertility problems, which is almost three times more than those who were unmarried. Similarly, the odds ratio for respondents being of fertility age (18-45) is 1.5. Females, employed and individuals residing outside county capitals were more likely to be acquainted with infertile couples. In this respect there is also a south-north gradient as southern respondents were more likely than central region respondents, who in turn were more likely than northern respondents, to be acquainted with infertile couples. None of the tested models show that income or the self-reported socio-economic condition had a significant impact on the dependent variable. Overall, about 5% of the variability in the dependent variables is explained by the explanatory variables. However, the hypothesis that none of the variables has an impact on respondents' acquaintance with couples with infertility problems is strongly rejected (p-value of less than 0.0001).

Table 6.5 presents the logistic regression models used to understand the main factors explaining respondents' knowledge of IVF. The binary variable (yes/no) is the answer to the question *"Did you know about IVF before* (this survey)"? Here the overall variability explained by the models is higher (about 7%) although fewer variables are statistically significant. In all three models, the level of education is highly significant (the higher the level of education, the higher the chance that the respondent knows of IVF). Gender and being married are also associated with the knowledge of IVF as females and married individuals were more likely to know about it. In contrast with the previous set of models (focused on acquaintance with infertile couples), age (in a quadratic definition) rather than fertility age (binary variable for individuals in their fertility age) is positively associated with the knowledge of IVF.

Table 6.4. Logistic regression: Acquaintance with infertile couples and socio-economic characteristics (dependent variable: *Are you acquainted with couples who desire to have babies and cannot have them?* Yes (1) and No (0)).

	Full dataset: All dependent variables included		Full dataset: Stepwise regression model (5% significance level)	
Number of observations	5,739		5,739	
Likelihood Ratio χ^2 (p-value)	420.56 (<0.0001)		391.78 (<0.0001)	
Pseudo R^2	0.0537		0.0500	
Count R^2	0.629		0.629	
	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)
Age				
B_1(age)	1.089	4.79 (<0.001)	1.087	4.72
B_2(age^2)	0.999	-4.59(<0.001)	0.999	-4.50
Fertility age*	1.6635	4.87 (<0.001)	1.647	4.96 (<0.001)
Gender	1.310	4.51 (<0.001)	1.128	4.55 (<0.001)
Education**	1.174	4.20 (<0.001)	1.176	4.26 (<0.001)
Employed (yes=1, no=0)	1.239	3.08 (0.002)	1.245	3.17 <0.002)
North (yes=1, no=0)	0.855	-1.95 (0.052)		
South (yes=1, no=0)	1.281	3.12 (0.002)	1.402	5.93 (<0.001)
County town (yes=1, no=0)	0.795	-3.99 (<0.001)	0.805	-3.77 (<0.001)
Married*** (yes=1, no=0)	1.862	6.73 (<0.001)	2.4688	6.86 (<0.001)
Number of children	0.957	-1.50 (0.134)		
Self-reported socio-economic status or monthly net income	0.985	-0.67 (0.500)		

* from age 18 to age 44

** Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

*** Married include individuals who live with their partners

Table 6.5. Logistic regression: Knowledge of In-Vitro-Fertilisation and socio-economic characteristics (dependent variable: *Did you know In-Vitro-Fertilisation before?* Yes (1), No (0))

	Full dataset: all dependent variables included		Full dataset: stepwise regression model (5% significance level)		Income dataset: stepwise regression model (5% significance level)	
Number of observations	5,739		5,739		4,806	
Likelihood Ratio χ^2 (p-value)	365.89 (p<0.0001)		362.94 (p<0.0001)		257.35	
Pseudo R^2	0.0803		0.0796		0.0719	
Count R^2	0.865		0.865		0.8774	
	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)
Age						
B_1(age)	1.102	4.29 (<0.001)	1.106	4.93 (<0.001)	1.159	9.07 (<0.001)
B_2(age^2)	0.999	-3.70(<0.001)	0.999	-4.34(<0.001)	0.998	-7.97(<0.001)
Fertility age*	1.120	0.73 (0.532)				
Gender	1.342	3.43 (0.001)	1.309	3.36(<0.001)	1.336	3.18 (0.001)
Education**	2.330	13.67 (<0.001)	2.369	14.17(<0.001)	1.943	9.75 (<0.001)
Employed (yes=1, no=0)	1.095	0.91 (0.362)				
North (yes=1, no=0)	1.080	0.66 (0.509)				
South (yes=1, no=0)	0.840	-1.53 (0.125)	0.776	-3.15 (0.002)		
County town (yes=1, no=0)	1.001	0.01 (0.990)				
Married*** (yes=1, no=0)	1.265	1.80 (<0.073)	1.6861	2.05(<0.041)		
Number of children	0.966	-0.81 (0.418)				
Self-reported socio-economic status or monthly net income (****)	1.129	3.80 (<0.001)	1.1229	3.71 (<0.001)	1.168	4.50 (<0.001)

* from age 18 to age 44

** Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

*** Married include individuals who live with their partner

**** Net income in Million Italian Liras (1€ = 1936.27 Liras)

About 86% of the sample answered “yes” to the “*Did you know about IVF before?*” question. For these respondents a multiple regression model is used to identify those variables which have an effect on the self-reported level of knowledge of IVF (table 6.6). Models based on log and exponential transformations of the dependent variable were tested. The models without any transofrmation consistently outperformed the others. However, our final the models explain only 4/5% of total variability. Again, education gender and marital status are the most significant variables. Males, less educated and unmarried individuals tend to have a more limited knowledge about IVF.

The last set of models investigates the answer to the question “”*Are you acquainted with someone who used In-Vitro-Fertilisation?*” (table 6.7). Logistic regression is used here to explain the impact of socio-economic characteristics of respondents on the variability of the yes/no answer to this question. The three logistic regressions present low explanatory power (about 4%). In both the models that resulted from the backward stepwise selection being of fertility age, being female, having a higher educational level, being married, being employed and living in the central and southern part of the country are positively associated with the probability of respondent's acquaintance with someone who used IVF. Neither socio-economic status nor income has any significant effect on the dependent variable.

Table 6.6. Multiple regression: self reported level of Knowledge of In-Vitro-Fertilisation and socio-economic characteristics (dependent variable: *You would state that your level of knowledge of In-Vitro-Fertilisation is*: very limited (1), limited (2), intermediate (3), deep (4), very deep (5))

	Full dataset: all dependent variables included		Full dataset: stepwise regression model 5% significance level))	
Number of observations	4,960		4,960	
Fisher test	(12, 4,947) 19.65		(6, 4,953) 38.55	
P-value (p> F)	<0.0001		<0.0001	
R^2	0.0455		0.0446	
	Coefficient	t-student (p-value)	Coefficient	t-student (p-value)
Age				
B_1(age)	0.01695	2.20 (<0.028)	0.01644	2.17 (0.030)
B_2(age^2)	0.00022	-2.60 (0.009)	0.00023	-2.75 (<0.006)
Fertility age*	0.0435	0.97 (0.332)		
Gender	0.1137	4.35 (<0.001)	0.1126	4.32 (<0.001)
Education**	0.1943	11.45<0.001)	0.1972	11.79 (<0.001)
Employed (yes=1, no=0)	-0.0569	-1.86 (0.062)	-0.0628-	2.10 (0.036)
North (yes=1, no=0)	-0.0166	-0.47 (0.637)		
South (yes=1, no=0)	-0.0011	-0.03 (0.996)		
County town (yes=1, no=0)	0.0250	0.99 (0.324)		
Married*** (yes=1, no=0)	0.1168	2.91 (<0.004)	0.1169	2.97 (0.003)
Number of children	0.0191	1.48 (0.139)		
Self-reported socio-economic status	0.0043	0.44 (0.663)		
Constant	1.3830	8.18 (<0.001)	1.4948	26.33 (<0.001)

* from age 18 to age 44

** Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

*** Married include individuals who live with their partner

Table 6.7. Logistic regression: Acquaintance with someone who used In-Vitro-Fertilisation and socio-economic characteristics (dependent variable: *Are you acquaint with someone who used In-Vitro-Fertilisation?* Yes (1), No (0))

	Full dataset: all dependent variables included		Full dataset: stepwise regression model (5% significance level)	
Number of observations	5,739		5,739	
Likelihood Ratio χ^2 (p-value)	211.19		203.25	
Pseudo R^2	0.0405		0.0394	
Count R^2	0.847		0.847	
	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)
Age				
B_1(age)	1.0800	2.91 (0.004)	1.07591	2.72 (0.007)
B_2(age^2)	0.99901	-3.18(0.001)	0.9989	-3.37 (0.001)
Fertility age*	1.2524	1.63 (0.104)		
Gender	1.6592	6.25 (<0.001)	1.7831	6.24 (<0.001)
Education**	1.3080	5.16 (<0.001)	1.6576	5.31 (<0.001)
Employed (yes=1, no=0)	0.2112	2.05 (<0.040)	1.2180	2.14 (<0.033)
North (yes=1, no=0)	0.8437	-1.54 (0.123)	0.7880	-3.00 (0.003)
South (yes=1, no=0)	1.1274	1.13 (0.260)		
County town (yes=1, no=0)	0.9071	-1.25 (0.210)		
Married*** (yes=1, no=0)	1.6634	3.96 (<0.001)	1.7880	4.51 (<0.001)
Number of children	0.9450	-1.43 (0.154)		
Self-reported socio-economic status or monthly net income	1.0041	0.13 (0.893)		

* from age 18 to age 44

** Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

*** Married include individuals who live with their partner

Results from these regression analyses provide evidence in favour of the validity of the survey and are fairly consistent with expectations. Females are closer to the issues of infertility and ARTs than men. This is probably due to the fact that women, in addition to the issues related to parenthood, feel that they have a more important role in procreation. Two sets of regressions refer to acquaintance of the respondents with someone affected by infertility and/or who used IVF. For these questions respondents are expected to report something stemming from their social life (being in touch with other people, being reported such facts by others, being in social environments where infertility and the use of IVF are discussed). In the other two sets of regressions, instead, the focus is on respondents "knowledge" of IVF that is the knowledge that respondents have of a new technology raising difficult scientific and ethical issues. In this case it may be expected that both cultural and social factors direct affect respondents' knowledge. Therefore it can be also expected that, compared to the former, the latter models (those focused on "knowledge"): i) present more overall explanatory power; and ii) have education playing a more significant role. Data support both these claims. The pseudo R^2 for the models on knowledge of IVF (about 0.07) is higher than those for acquaintance with infertility couples (0.05) or acquaintance with someone who used IVF (0.04).

Education is a highly significant explanatory variable in all models. However, the magnitude of the odd ratios varies significantly across the models (table 6.4, 6.5 and 6.7). Table 6.8 present a direct comparison of the effects of education on the dependent variables investigated in the 3 logistic regressions. For each model the fitted values of the regression models are calculated on the basis of the estimated coefficients. The table reports means and standard deviations of the probabilities for each of the education groups. In the model where the dependent variable is the respondent's acquaintance with infertile couples, the least educated people had a probability of 0.34 of having acquaintance with an infertile couple compared to a probability of 0.52 of the most educated people. Similarly, acquaintance with someone who used IVF is less likely in the least educated respondents (0.09) than in the most educated ones (0.23). The magnitude of these differences (0.18 and 0.14, respectively) is relevant but smaller than the difference between the two education groups with respect to knowledge of IVF (0.73 and 0.97 for a difference of 0.24). Indeed, the odds ratio for the education variable in the logistic regression for knowledge of IVF is 2.12.

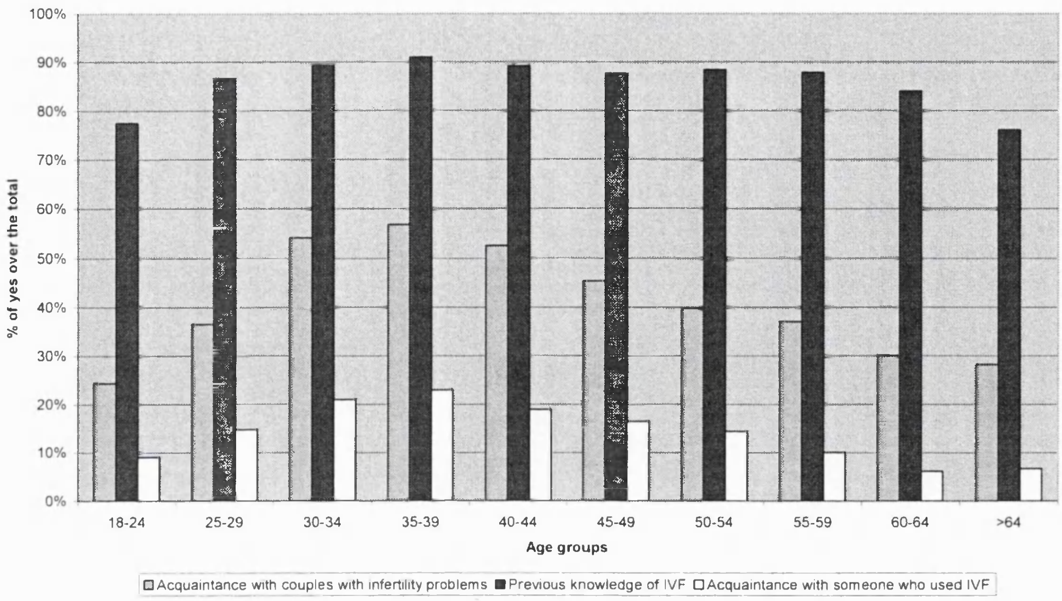
Table 6.8. Comparison between the coefficients (ß) for education on three logistic regression models on the knowledge of infertility and IVF (backward stepwise logistic regression on the full dataset of 5,739 respondents).

	Number of respondents	Mean probability (standard deviation) fitted by the stepwise regression model		
		Acquaintance with infertile couples	Acquaintance with someone who used IVF	Knowledge of IVF
Less than 8 year of school	609	0.335 (0.093)	0.090 (0.030)	0.730 (0.059)
Between 8 and 13 years of school	2068	0.406 (0.115)	0.134 (0.049)	0.824 (0.072)
High school diploma	2577	0.446 (0.130	0.173 (0.065)	0.908 (0.043)
University degree	485	0.520 (0.114)	0.232 (0.074)	0.970 (0.014)

Age was expected to be associated with the answers to all these questions. We used the age of the respondents in two ways. The first was to use a quadratic expression of age as a regressor. Both the coefficients (of the linear and of the quadratic term) are significant in all models. Basically, young and old persons are less knowledgeable of (and close to) IVF issues than peple in the central part of their life. ; The second way to use age as independent variable was to introduce a dummy variable to identify individuals presumed to be of "fertility age" (from 18 to 45). This dummy is significant in two models. Respondents of fertility age are more likely to be acquainted with infertile couples and with people who used IVF. This is probably the effect of friendship; people tend to have stronger friendship relationships with people of similar age and thus it is more likely that respondents in the age of fertility know infertile couples and/or people who used IVF.

For all the questions of this part of the survey it can be observed a curvilinear relation with age (figure 6.1). The youngest respondents are acquainted with few people with infertility problems and who used IVF; also, they are not very knowledgeable about IVF. The peak is in the 35-39 age bracket; then older patients tend to be less acquainted with infertile couples and with individuals who used IVF; furthermore, people in this age bracket also appear to know less about IVF. These relationships make sense as it is reasonable to assume that these are issues for which people in their late 30s are the most sensitive. Women, and to a lesser extent men, between 35 and 45 are the main users of IVF and thus are more likely to be informed about its use and about other several related issues.

Figure 6.1 The curvilinear relation between age and knowledge of infertility and IVF



6.4. Hypothetical personal use of IVF

A specific question asked respondents if they would have personally tried IVF in a hypothetical situation of infertility. Almost 20% of the sample (1,119 respondents) answered "do not know" (table 6.9). About 34% answered that they would have not tried IVF, while 47% answered that they would have tried. Among "yes" answers the percentage of "probably" (39.1%) is much higher than that of "definitely" (14.6%), while in the “no” answers the two percentages are very similar (17.4% “definitely not” and 17.1% “probably not”). If the “do not know” answers are excluded from the analysis, individuals willing to try IVF are the majority of the sample (57.2%). Very minimal differences between males and females appear in the answers to this question.

Table 6.9. Respondents' attitude towards IVF in a hypothetical condition of infertility. Answers to the question: "Imagine that you have been married for a few years and you have been unsuccessfully trying to have a baby. Imagine you are told to use IVF and that it has a 30% chance to be successful. Would you personally try IVF?"

	Including "do not know"						Excluding "do not know"					
	#			%			#			%		
	All	Females	Males	All	Males	All	All	Females	Males	All	Males	All
Do not know	1,119	537	582	19.5	18.2	20.9						
No, definitely not	998	539	459	17.4	18.3	16.4	998	539	459	21.6	22.4	20.8
No, probably not	980	487	493	17.1	16.6	17.7	980	487	493	21.2	20.2	22.3
Yes, probably	1,805	924	881	31.4	31.4	31.6	1,805	924	881	39.1	38.3	39.9
Yes, definitely	837	460	377	14.6	15.6	13.5	837	460	377	18.1	19.1	17.1
Total	5,739	2,947	2,792	100.0			4,620	2,410	2,210	100.0	100.0	100.0

In the logistic regression models approximately 5% of variability is explained by the dependent variables (table 6.10). In the full data set being younger, being employed, being married, having less children, being more educated and living in northern regions and in county towns are positively associated with the willingness to try IVF. However, none of these variables show odds ratios of relevant magnitude (they range from 0.9 to 1.3). The only relevant exception is education because the coefficient is 1.17 for each of the four levels of education and the p-value of the coefficient is less than 0.001. Nevertheless, the most relevant regressors (highest odds ratios and Z values) are the answers to the questions presented in the previous section. Respondents who are acquainted with infertile couples (OR = 1.42), who knew IVF before (OR = 2.01) and who are acquainted with someone who used IVF (OR = 1.7) are more likely to provide a positive answer to the question "Would you personally try IVF?"

Results of the logistic regression on the income data set are very similar. For all but two independent variables coefficients are similar and statistically significant. In the income data set the variable "number of children", which was negatively associated with the willingness to try IVF in the full data set, is excluded by the stepwise procedure and the variable "income" is included (odds ratio greater than 1).

The explanatory power of multiple regression models is still low ($R^2 = 0.07$) and their coefficients are consistent with those obtained in the logistic analyses. In both dataset age (-), education (+), employment (+), living in a county town (+) and being married (+) are associated with the willingness to try IVF measured on a scale from 1 to 5. As in the logistic regressions, respondents who reported to be informed about IVF and acquainted with people who where infertile and/or used IVF are significantly more likely to be in favour of using IVF in case of need. Moreover, if regressions are run with the income dataset income shows a positive coefficient (p-value <0.001) and the number of children is excluded from the model by the stepwise procedure.

Table 6.10. Logistic regression: use of IVF in case of hypothetical infertility (dependent variable: "Imagine that you have been married for a few years and you have been unsuccessfully trying to have a baby. Imagine you are told to use IVF and that it has a 30% chance to be successful. Would you personally try IVF?" Yes (1), No (0))*

	Full dataset: all dependent variables included		Full dataset: stepwise regression model (5% significance level)		Income dataset: stepwise regression model (5% significance level)	
Number of observations	4,620		4,620		3,885	
Likelihood Ratio χ^2 (p-value)	328.31 (<0.0001)		323.26 (<0.0001)		255.11 (<0.0001)	
Pseudo R^2	0.05		0.0512		0.0482	
Count R^2	0.623		0.622		0.626	
	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)
Age						
B_1(age)	1.0551	2.72 (0.007)	1.0587	4.10 (<0.001)	1.0600	3.82 (<0.001)
B_2(age^2)	0.9992	-3.57(<0.001)	0.9992	-5.08 (<0.001)	0.9992	-4.66(<0.001)
Fertility age**	1.0775	0.66 (0.506)				
Gender (female=1, male=0)	0.9693	-0.47 (0.640)				
Education***	1.1689	3.64 (<0.001)	1.1894	4.16 (<0.001)	1.1903	3.72 (<0.001)
Employed (yes=1, no=0)	1.0547	0.69 (0.488)		2.04 (0.041)		
North (yes=1, no=0)	0.9595	-0.46 (0.643)				
South (yes=1, no=0)	1.0781	0.85 (0.398)				
County town (yes=1, no=0)	1.2994	4.09 (<0.001)	1.2918	3.99 (<0.001)	1.2418	3.14 (0.002)
Married**** (yes=1, no=0)	1.0108	0.10 (0.917)				
Number of children	0.9036	-3.05 (0.002)	0.9207	-2.94 (0.003)		
Self-reported socio-economic status or monthly net income *****	1.0251	0.98 (0.357)			1.0788	3.31 (0.001)
Acquaintance with infertile couples (yes=1, no=0)	1.4064	5.11 (<0.001)	1.4227	5.34 (<0.000)	1.4025	4.67 (<0.001)
Knowledge of IVF (yes=1, no=0)	1.9739	6.43 (<0.001)	2.0000	6.40 (<0.000)	1.9830	5.68 (<0.001)
Acquaintance with someone who used IVF (yes=1, no=0)	1.7024	5.86 (<0.001)	1.7016	5.87 (<0.000)	1.6468	5.07 (<0.001)

* "do not know" answers excluded from the analysis
** from age 18 to age 44
*** Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)
**** Married includes individuals who live with their partner
***** Net income in Million Italian Liras (1 € = 1936.27 Liras)

Table 6.11. Multiple regression: use of IVF in case of hypothetical infertility (dependent variable: "Imagine that you have been married for a few years and you have been unsuccessfully trying to have a baby. Imagine you are told to use IVF and that it has a 30% chance to be successful. Would you personally try IVF?")

	Full dataset: All dependent variables included		Full dataset: stepwise regression model (5% significance level))		Income dataset: stepwise regression model (5% significance level))	
Number of observations	5,739		5,739		4,806	
Fisher test	(14; 5,724)		(8; 5,730)		(9; 4,796)	
P-value (p> F)	<0.0001		<0.0001		<0.0001	
R^2	0.0662		0.0666		0.0664	
Dependent variable: yes definitely (5), yes probably (4), do not know (3), no probably (2), no definitely (1)						
	Coefficient	t-student (p-value)	Coefficient	t-student (p-value)	Coefficient	t-student (p-value)
Age						
B_1(age)	0.01962	2.72 (0.007)	0.02762	3.78 (<0.001)	0.02739	3.43 (<0.001)
B_2(age^2)	-0.00034	-3.57(<0.001)	-0.00043	-5.01 (<0.001)	-0.00043	-4.57 (<0.001)
Fertility age*	0.0387	0.62 (0.533)				
Gender	-0.0184	-0.11 (0.915)				
Education**	0.0784	3.57 (0.000)	0.0888	3.84 (<0.001)	0.0875	3.37 (0.001)
Employed (yes=1, no=0)	0.0522	2.46 (0.014)				
North (yes=1, no=0)	-0.0410	-0.83 (0.937)				
South (yes=1, no=0)	0.0028	4.18 (<0.001)				
County town (yes=1, no=0)	0.1465	2.73 (0.006)	0.1436	4.13 (<0.001)	0.1037	2.73 (0.006)
Married*** (yes=1, no=0)	0.0425	0.76 (0.446)		-		
Number of children	-0.0511	-2.83 (0.005)	-0.4445	-2.52 (0.012)	0.0384	-2.01 (0.036)
Self-reported socio-economic status or monthly net income****	0.01683	1.22 (0.221)			0.0542	4.32 <(0.001)
Acquaintance with infertile couples (yes=1, no=0)	0.2173	5.95 (<0.001)	0.2243	6.06 (<0.001)	0.2154	5.31 (<0.001)
Knowledge of IVF (yes=1, no=0)	0.3236	6.30 (<0.001)	0.3329	6.30 (<0.001)	0.3250	5.58 (<0.001)
Acquaintance with someone who used IVF (yes=1, no=0)	0.3936	7.88 (<0.001)	0.3947	7.93 (<0.001)	0.3653	6.74 (<0.001)
Constant	2.1929	9.66 (<0.00)	2.1207	13.52 (<0.001)	1.9736	11.44 (<0.001)

* from age 18 to age 44

** Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

*** Married includes individuals who live with their partner; **** Net income in Million Italian Liras (1€ = 1936.27 Liras)

Coefficients estimated in the regression models were used to predict the probability that the respondent was willing to try IVF in case of infertility for specific values of the dependent variables (e.g. income, education, age, etc.). Predictions were generated by the income data set and by the logistic regression model specified by the backward stepwise procedure. The predicted probability that an individual of the sample is willing to try IVF is 0.57. Although their coefficients are statistically significant, some variables do not predict substantial differences (table 6.12). Marital status and place of residence are not relevant predictors. The role of employment status is more important; while unemployed respondents have a probability of 0.51 those who are employed have a probability of 0.61. Education is the most important predictor. There is almost an absolute 30% difference between the most educated individuals, those who hold a University degree (probability = 0.67), and the least educated (probability = 0.39). Income is also important: for respondents in the low income brackets the probability of trying IVF is below 0.5, while for high-income respondents the probability ranges from 0.6 to 0.65. Overall, however, education is much more important than income in explaining the variability of the attitude towards the use of IVF.

The logistic regression model predicts that the highest probability of trying IVF can be found in the 30 year old respondents (probability = 0.62). As age increases, the probability of the willingness to try IVF declines this is up to the value of 0.36 for respondents who are over 64. However, the relationship between age and attitude towards IVF is not linear because respondents in their twenties show a lower probability than those in their thirties.

The probability of a "yes" answer to the question *"Would you try IVF?"* is higher if the respondents are acquainted with infertile couples (0.64), if they knew IVF before the survey (0.59) and if they are acquainted with someone who used IVF (0.72). Individuals who reported all three characteristics (i.e. being acquainted with infertile couples, knowing about IVF, and being acquainted with couples who tried IVF) show a probability of 0.74. Those who lack all the three characteristics have a probability of 0.34. These "soft" characteristics are by far the most important predictors of the attitude towards the use of IVF.

Table 6.12. Factors explaining willingness to try IVF: predictions from the logistic regression (Income dataset; dependent variable: *"Imagine that you have been married for a few years and you have been unsuccessfully trying to have a baby. Imagine you are told to use IVF and that it has a 30% chance to be successful. Would you personally try IVF?"*)

Respondents characteristics	# of observations	Expected probability of trying IVF	Respondents' characteristics	# of observations	Expected probability of trying IVF
Age			Monthly Family Income*		
18-24	714	0.56	700.000 (€ 362)	47	0.47
25-29	418	0.60	900.000 (€ 465)	64	0.49
30-34	546	0.62	1.250.000 (€ 646)	272	0.48
35-39	702	0.62	1.750.000 (€ 904)	803	0.53
40-44	640	0.60	2.250.000 (€ 1,162)	904	0.55
45-49	582	0.57	2.275.000 (€ 1,175)	652	0.57
50-54	518	0.53	3.250.000 (€ 1,678)	536	0.57
55-59	321	0.49	3.750.000 (€ 1,937)	545	0.61
60-64	186	0.43	4.500.000 (€ 2,324)	502	0.60
>64	179	0.36	5.500.000 (€ 2,841)	249	0.61
Education			6.500.000 (€ 3,357)	86	0.65
Less than 8 years of education	482	0.39	7.500.000 (€ 3,873)	48	0.60
Between 8 and 13 years of education	1,727	0.53	8.500.000 (€ 4,390)	98	0.65
High School Diploma (at least 13 years of education)	2,191	0.61			
University Degree	406	0.67			
Employed			Acquaintance with infertile couples		
Yes	2,671	0.61	Yes	2,049	0.64
No	2,135	0.51	No	2,757	0.51
County town			Knowledge of IVF		
Yes	1,929	0.59	Yes	4,216	0.59
No	2,877	0.55	No	590	0.36
Married			Acquaintance with someone who used IVF		
Yes	3,376	0.57	Yes	759	0.72
No	1,430	0.54	No	4,047	0.54

* Net of income taxes

6.5. Variables explaining willingness to pay for personal use of IVF and for a publicly funded programme

Three questions were framed to elicit willingness to pay for IVF. First, a question was presented to elicit how much respondents were willing to pay in the case they wanted to use IVF. Respondents were presented 6 values and for each of them they were asked to choose between "yes" and "no". Then, we administered the WTP question for a national programme providing public funding for IVF according to a referendum format. The take-it-or-leave-it question was presented with four different values by splitting the total sample into four sets. Each sub sample consisted of about 1,400 individuals. Finally, all respondents were requested to answer to a series of modified payment card WTP questions about the national programme.

We ran a set of regressions for each of the three WTP questions as dependent variables, using basic socio-demographic characteristics as independent variables. For the WTP for personal use of IVF we censored data at both extremes values presented in the modified payment card. As linear regression models may not be appropriate we also used a doubled censored *Tobit* model, with censoring at the lower (€ 0) and upper (€ 25,823) values proposed to respondents The log transformation of stated WTP values presented higher pseudo R^2 (about 0.03 versus 0.01) and higher likelihood ratios than untransformed WTP values. Consequently, we used the log specification to report regression results (table 6.13 and 6.14).

The model based on the income data set presents the best goodness of fit values and the expected signs for the independent variables. Income and education are positively associated to the stated WTP for personal use of IVF. On the contrary, older and northern respondents present lower WTP. In this survey residing in a northern region is regularly associated with a weaker preference for IVF; likely, this is due to social characteristics and traditions of northern regions differentiating them from the central and, especially, the southern ones. Both coefficients of age are significant, suggesting that the relationship between WTP and age is indeed quadratic, with WTP relatively low and increasing in the twenties and thirties, at its peak in the early fourtys and decreasing since then.

Table 6.13. Tobit regression models of willingness to pay for per personal use of IVF in case of infertility: modified payment card data (dependent variable: 6 values ranging from It. Liras 0 to It. Liras 50.00.000 (€ 25,823); dependent variable = log(WTP+1) and censored at 0 and log (50.000). WTP data in thousand Italian Liras.

	Full dataset: all dependent variables included		Full dataset: stepwise regression model (5% significance level)		Income dataset: stepwise regression model (5% significance level)	
Number of observations	5,739		5,739		4,806	
Likelihood Ratio χ^2 (p-value)	226.05 (<0.0001)		2.16.76 (<0.0001)		1009.56 (<0.0001)	
Pseudo R^2	0.0093		0.0089		0.010	
	2,612 left-censored observations; 500 right-censored observations		3,174 left-censored observations; 500 right-censored observations		2,623 left-censored observations 422 right-censored observations	
	Coefficient	t-student (p-value)	Coefficient	t-student (p-value)	Coefficient	t-student (p-value)
Age						
B_1(age)	0.29030	3.63 (<0.001)	0.3952	3.63 (<0.001)	0.37875	6.29 (<0.001)
B_2(age^2)	-0.00391	-4.39(<0.001)	-0.00520	-4.39(<0.001)	-0.000722	-7.18 (<0.001)
Fertility age (1= age from 18 to 44, 0 otherwise)	0.5457	1.16 (0.245)				
Gender (female=1, male=0)	0.2735	1.16 (0.316)				
Education*	1.2775	7.19 (<0.001)	1.3331	7.71 (<0.001)	1.0004	5.20 (<0.001)
Employed (yes=1, no=0)	0.4203	1.32 (0.188)				
North (yes=1, no=0)	-0.8288	-2.24 (<0.025)	-0.9765	-3.64 (<0.025)	-1.3177	-4.47 (<0.001)
South (yes=1, no=0)	0.2419	0.66 (0.508)				
County town (yes=1, no=0)	0.9016	3.41 (0.001)	-0.8459	3.23 (<0.001)		
Married** (yes=1, no=0)	0.6436	1.53 (<0.125)				
Number of children	-0.3795	-2.77 (0.006)	-0.4202	-3.11 (0.006)	-0.3621	-2.50 (0.012)
Self-reported socio-economic status or monthly net income***	0.16568	1.59 (0.111)			0.6106	6.41 (<0.001)
Constant	-6.7210	-3.90 (<0.001)	-6.885	-5.63 (<0.001)	-6.7455	-5.16 (<0.001)
Sigma	8.84		8.85		8.68	

* Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)
** Married includes individuals who live with their partner
*** Monthly net income reported in Million Italian Liras (1 Euro = 1,937 Italian Liras)

Table 6.14. Linear regression models of willingness to pay for per personal use of IVF in case of infertility: modified payment card data (dependent variable: 6 values ranging from It. Liras 0 to It. Liras 50,000,000 (€ 25,823); dependent variable = log(WTP+1) and censored at 0 and log (50.000.000). WTP data in thousand Italian Liras

	Full dataset: all dependent variables included		Full dataset: stepwise regression model (5% significance level)		Income dataset: stepwise regression model (5% significance level)	
Number of observations	5,739		5,739		4,806	
Likelihood Ratio χ^2 (p-value)	F(11, 5726) = 19.57 (Pr> F <0.0001)		F(11, 5726) = 19.57 (Pr> F <0.0001)		F(6, 4799) = 34.47 (Pr> F <0.0001)	
R^2	0.0394		0.0380		0.0413	
	Coefficient	t-student (p-value)	Coefficient	t-student (p-value)	Coefficient	t-student (p-value)
Age						
B_1(age)	0.12926	3.60 (<0.001)	0.18044	7.17 (<0.001)	0.16959	7.17 (<0.001)
B_2(age^2)	0.00166	-4.21(<0.001)	-0.00217	-7.22 (<0.001)	-0.00228	-7.22 (<0.001)
Fertility age (1= age from 18 to 44, 0 otherwise)	0.3993	1.85 (0.065)	0.4516	2.12 (<0.034)		
Gender (female=1, male=0)	0.0963	0.70 (0.486)				
Education*	0.4772	5.16 (<0.001)	0.6455	8.10 (<0.001)	0.5043	5.59 (<0.001)
Employed (yes=1, no=0)	0.2096	1.28 (0.201)				
North (yes=1, no=0)	-0.55106	-2.97 (<0.001)	-0.4399	-3.55 (<0.001)	-0.6087	-4.40 (<0.001)
South (yes=1, no=0)	0.1000	0.54 (0.672)				
County town (yes=1, no=0)	0.2685	2.80 (0.005)	0.3719	3.06 (0.002)		
Married** (yes=1, no=0)	0.2707	1.26 (<0.207)				
Number of children	-0.1450	-2.10 (0.036)	-0.1882	-3.01 (<0.003)	-0.17364	-2.56 (<0.010)
Self-reported socio-economic status or monthly net income***	0.2605	1.97 (0.049)			0.27951	6.26 (<0.001)
Constant	-0.502	0.64 (0.525)	-0.1169	-0.18 (<0.854)	0.5490	0.92 (<0.357)

* Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

** Married includes individuals who live with their partner

*** Monthly net income reported in Million Italian Liras (1 Euro = 1,937.27 Italian Liras)

Tobit and regression models present coefficients with the same sign and similar magnitude. However, their predictions should be taken cautiously, as the assumptions about the tails of the distribution significantly influences mean values. For example, in both models the income coefficient is significantly positive and suggests that about € 500 increase in income is associated to an increase by 30-60% of the WTP for private use of IVF. However, both sets of models greatly underestimate mean WTP (see later).

Table 6.15 reports the results of the three regression models using the answers to the TIOLI questions on WTP for a publicly funded programme (full data set with all possible independent variables, full data set and backward stepwise selection, and data set containing information about family income and backward stepwise selection). For all models, all but three variables are significant at the 5% level. The variable with the highest absolute Z score is the value of the bid (OR: 0.98; z= -8.8). As expected, the higher the value of the bid, the lower the probability of a "yes" answer. The other significant variables, retained by the stepwise procedure with both datasets, are age and number of children, with the negative sign, and education, employment status, living in a southern region, living in a county town and being married, with the positive sign. Although the explanatory power of the model is limited (R^2 around 0.03), its regressors have the expected signs and are generally consistent with regression models tested in the previous sections.

The ordered logistic regressions are expected to capture the intensity of the preference of respondents (table 6.15). Virtually all the regressors identified in the logit binary model are also significant in the ordered logistic regressions. The only marginal difference concerns the two dummy variables for the three macro geographic areas (north, centre and south). While in the binary logit the "South" dummy is significant with the positive sign (people residing in the southern regions were more likely to be willing to pay for the IVF programme), in the ordered logit the dummy for the northern regions is significantly negative.

Table 6.15. Logistic regression of the willingness to pay for a publicly funded IVF programme for infertile couples: referendum data (dependent variable: Yes definitely or probably (=1) and No definitely or probably (=0) to the question *"Would you vote in favour of funding IVF with public money if you were asked an annual payment, that is a tax, of Italian £ X?""*; bids = It Liras 5,000 (€ 2.6), 10,000 (€ 5.2), 20,000 (€ 10.3) and 50,000 (€ 25.8)).

	Full dataset: all dependent variables included		Full dataset: stepwise regression model (5% significance level)		Income dataset: stepwise regression model (5% significance level)	
Number of observations	4,838		4,838		4,071	
Likelihood Ratio χ^2 (p-value)	214.01 (<0.0001)		212.31		187.41	
Pseudo R^2	0.0327		0.0324		0.0341	
	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)
Bid*	0.9852	-8.85 (<0.001)	0.9852	-8.85 (<0.001)	0.9840	-8.84 (<0.001)
Age						
B_1(age)	1.052	3.60 (<0.001)	1.060	4.17 (<0.001)	1.0470	3.22 (0.001)
B_2(age^2)	0.99928	-4.21(<0.001)	-0.00327	-4.52 (<0.001)	-0.99931	-4.03 (<0.001)
Fertility age (1= age from 18 to 44, 0 otherwise)	1.1235	1.07 (0.284)				
Gender (female=1, male=0)	0.9950	-0.08 (0.936)				
Education**	1.1691	3.80 (<0.001)	1.1705	3.83 (<0.001)	1.0978	2.01 (0.045)
Employed (yes=1, no=0)	1.1629	2.18 (0.029)	1.1681	2.37 (0.018)	1.1745	2.23 (0.026)
North (yes=1, no=0)	0.9354	-0.77 (0.441)				
South (yes=1, no=0)	1.0979	1.08 (0.280)	1.1515	2.25 (0.024)	1.1813	2.34 (0.019)
County town (yes=1, no=0)	1.1404	2.12 (0.034)	1.1515	2.25 (0.024)	1.1952	2.65 (0.008)
Married*** (yes=1, no=0)	1.5824	5.51 (0.017)	1.1450	2.20 (0.028)	1.5069	4.56 (<0.001)
Number of children	0.9150	-2.76 (0.006)	0.9134	-2.83 (0.011)	0.9008	-2.96 (0.003)
Self-reported socio-economic status or monthly net income****	1.0661	2.63 (0.09)	1.0657	2.62 (0.009)	1.0903	3.70 (<0.001)

* (It Liras 5,000 (€ 2.6), 10,000 (€ 5.2), 20,000 (€ 10.3) and 50,000 (€ 25.8))

** Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

*** Married includes individuals who live with their partner

**** Monthly net income reported in Million It. Liras (1 Euro = 1,936.27 Italian Liras)

Table 6.16. Ordered Logistic regression of the willingness to pay for a publicly funded IVF programme for infertile couples: referendum data (dependent variable: no, definitely not (1); no, probably not (2), do not know (3); yes, probably (4); yes, definitely (5) to the question *"Would you vote in favour of funding IVF with public money if you were asked an annual payment, that is a tax, of Italian £ X?""*: Bids = It. Liras 5,000 (€ 2.6), 10,000 (€ 5.2), 20,000 (€ 10.3) and 50,000 (€ 25.8)).

	Full dataset: all dependent variables included		Full dataset: stepwise regression model (5% significance level)		Income dataset: stepwise regression model (5% significance level)	
Number of observations	5,739		5,739		4,806	
Likelihood Ratio χ^2 (p-value)	271.39 (<0.0001)		268.36		238,75	
Pseudo R^2	0.0149		0.0147		0.0157	
	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)	Odds Ratio	Z (p-value)
Age						
B_1(age)	1.035	3.83 (<0.001)	1.045	3.97 (<0.001)	1.0470	3.62 (0.001)
B_2(age^2)	0.99928	-4.56(<0.001)	-0.99971	-4.02 (<0.001)	-0.99954	-3.95 (<0.001)
Fertility age (1= age from 18 to 44, 0 otherwise)	1.1526	1.65 (0.100)				
Gender (female=1, male=0)	1.0183	0.36 (0.715)				
Education**	1.1601	4.55 (<0.001)	1.1633	4.66 (<0.001)	1.1121	2.91 (<0.004)
Employed (yes=1, no=0)	1.1876	3.14 (0.002)	1.1843	3.27 (<0.001)	1.1645	2.67 (<0.008)
North (yes=1, no=0)	0.8843	-1.83 (0.068)	0.8727	-2.75 (0.006)	0.8700	-2.53 (0.011)
South (yes=1, no=0)	1.0214	0.31 (0.754)				
County town (yes=1, no=0)	1.1380	2.64 (0.068)	1.1339	2.60 (0.009)	1.1354	2.40 (0.016)
Married*** (yes=1, no=0)	1.4582	5.78 (<0.001)	1.4911	6.25 (<0.001)	1.4610	5.39 (<0.001)
Number of children	0.9198	-3.33 (0.001)	0.9170	-3.48 (<0.001)	0.9175	-3.17 (<0.002)
Self-reported socio-economic status or monthly net income****	1.0487	2.50 (0.012)	1.0472	2.43 (0.015)	1.0856	4.59 (<0.001)

* (It. Liras 5,000 (€ 2.6), 10,000 (€ 5.2), 20,000 (€ 10.3) and 50,000 (€ 25.8))

** Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

*** Married includes individuals who live with their partner

**** Monthly net income reported in Million Italian Liras (1 Euro = 1,936.27 Italian liras)

Regression analysis models of the MPC data support the theoretical validity of the WTP elicitation method. In all models education is a significant predictor of the magnitude of willingness to pay for the IVF programme. In the full dataset stepwise regression model one class difference in education (respondents are classified according to 4 classes) is associated with an expected increase in willingness to pay for the programme of It. Liras 9,600 (€ 5) (table 6.16). The expected difference in the WTP between the least educated (less that 8 years of education) and the most educated (university degree) is almost € 15. In the analysis performed on the sub dataset reporting income values the coefficient of education is smaller (6.6 versus 9.6). However, income looks a better explanatory variable than the socio-economic status. For an increase of the net monthly income of € 500 it is expected an increase in WTP of about € 2.50, holding all the other variables constant. Married individuals (including those living with their partner) have higher WTP for the IVF programme ranging from € 6 to € 8 depending on the model specification. Unemployed individuals (including retired people) and those residing in northern regions present lower WTP for the IVF programme than those employed and residing in central and southern regions. Although the Likelihood Ratio tests suggest that we can reject the hypothesis that all coefficients are 0, it should be noted that the explained variability of the models is low (Pseudo R^2=0.03).

Table 6.17. Tobit regression models of willingness to pay for a publicly funded IVF programme for infertile couples: modified payment card data (dependent variable: 8 values ranging from It. Liras 2,000 (Euro 1) to It. Liras 200,000 (€ 103.3); dependent variable censored at It. Liras 0 and Italian Liras 200,000).

	Full dataset: all dependent variables included		Full dataset: stepwise regression model (5% significance level)		Income dataset: stepwise regression model (5% significance level)	
Number of observations	5,739		5,739		4,806	
Likelihood Ratio χ^2 (p-value)	136.55 (<0.0001)		120.20		116.56	
Pseudo R^2	0.032		0.031		0.033	
	Coeff.	t-student (p-value)	Coeff.	t-student (p-value)	Coeff.	t-student (p-value)
Age						
B_1(age)	321.56	0.37 (0.718)	439.35			
B_2(age^2)	-7.22	-0.73 (<0.464)	-10.73			
Fertility age (1= age from 18 to 44, 0 otherwise)	10.4010	1.97 (0.049)	16.86	5.13 (<0.001)		
Gender (female=1, male=0)	5.7430	1.89 (0.059)				
Education*	9.0150	4.54 (<0.001)	9.5695	4.89 (<0.001)	6.5966	2.99 (0.003)
Employed (yes=1, no=0)	10.0375	3.00 (0.003)	8.0386	2.54 (0.011)	8.5498	2.47 (0.013)
North (yes=1, no=0)	-14.9271	-3.60 (<0.001)	-15.4242	-5.13 (<0.001)	-19.0558	-5.74 (<0.001)
South (yes=1, no=0)	1.9233	0.47 (0.639)				
County town (yes=1, no=0)	7.6121	2.56 (0.011)	7.5348	2.56 (0.011)		
Married** (yes=1, no=0)	14.0424	3.50 (0.001)	12.4714	3.56 (<0.001)	16.2373	3.78 (<0.001)
Number of children	-2.1108	-1.37 (0.169)				
Self-reported socio-economic status or monthly net income***	2.3840	2.04 (0.041)	2.3771	2.04 (0.042)	4.8999	4.56 (<0.001)
Constant	-15.7656	-1.15 (0.250)	-31.1966	-4.51 (<0.001)		
Sigma	101.487		100.786		100.076	

* Education is classified in 4 classes: less than 8 years of education (1), between 8 and 13 years (2), high school diploma (3) and university degree (4)

** Married includes individuals who live with their partner

*** Monthly net income reported in Million Italian Liras (1 Euro = 1,936.27 Italian Liras).

Chapter 7

A cost-benefit analysis of In-Vitro-Fertilisation

7.1. Introduction

In this chapter we present the analysis of benefits and costs data that we obtained from the empirical study. The rationale of methodological choices and details about the data are presented in Chapters 4 and 5. Here, we first analyse the answers to the WTP questions and then provide estimates of mean WTP for hypothetical personal use of IVF (section 7.2) and for an Italian national programme providing IVF to Italian infertile couples who desire to have a baby (sections 7.3-7.7). WTP data collected according to two elicitation methods (take-it-or-leave-it and a modified version of payment card) are used to make some validity checks (section 7.8) and comparisons (section 7.9).

Section 7.10 presents the data obtained in the cost analysis conducted in two Italian hospitals. Results presented in the previous sections are then used to estimate the incremental cost-effectiveness ratio of IVF and to present a cost-benefit analysis of a national programme providing IVF (7.11 and 7.12). Overall, results show that providing IVF to infertile couples should imply a welfare gain.

7.2. Willingness to pay for hypothetical personal use of IVF

A specific question was framed to elicit how much respondents were willing to pay in case they wanted to use IVF. Respondents were presented 6 values and for each of them they were asked to choose between "yes" and "no". As detailed in Chapter 5, two main assumptions can be made about the rationality of respondents (potential users of IVF). The first assumption, which reflects standard economic theory, is that respondents are not expected to answer "yes" at any given amount if for lower amounts they have answered "no". The second assumption accepts that respondent can violate the assumption, but only once.

Under the first assumption, there is an inconsistency when a respondent states "yes" to a value that it is higher than a value for which he/she stated "no". According to this definition of inconsistency, for the total sample of 5,739 individuals there are 504 inconsistencies to this part of the questionnaire (table 7.1). As each respondent was presented 6 bids, the maximum possible number of inconsistencies is 34,434. Therefore, only 1.5% of the answers can be deemed invalid under this rational assumption. For several respondents there were more than one inconsistency and for few of them there is a pattern suggesting protest answers (see below). It is calculated that 207 respondents present at least one inconsistency to the answer to the willingness to pay for IVF. Accordingly, 3.6% of respondents provided an invalid answer to this question if the standard assumption of the positive correlation between price and willingness to pay is assumed.

If it is assumed that an answer is inconsistent only if the "normality" rule is violated twice, very few answers are void. Sixteen answers meet these criteria (0.3% of the sample). These answers are null even if a wide concept of rationality is assumed; it is also very likely that they express protest.

Overall this check corroborates the validity of this part of the questionnaire. If we impose a standard notion of rationality, 3.6% of respondents provided invalid answers; however, if the rationality assumption is relaxed only 0.3% respondents provided clearly void answers. Overall, it appears that this part of the questionnaire was comprehended and provoked very few protests.

Table 7.1. Willingness-to-pay for the use of IVF: analysis of inconsistent answers (under the assumption that the probability of stating the use of IVF increases as the prices increase).

	Willing to Pay £ 0 million	Willing to Pay £ 1 million	Willing to Pay £ 5 million	Willing to Pay £ 10 million	Willing to Pay £ 20 million	Willing to Pay £ 50 million	Total 1*	Total 2**
No Willing to Pay £ 0	X	51	29	33	26	24	163	5
No Willing to pay £ 1 mln		X	34	35	24	22	115	1
No Willing to pay £ 5 mln			X	50	35	33	118	5
No Willing to pay £ 10 mln				X	32	38	70	10
No Willing to pay £ 20 mln					X	38	38	38
No Willing to pay £ 50 mln						X		
Total							504	59

* Total 1 reports the number of inconsistent answers; the grand total is the total number of inconsistencies to the question on willingness to pay for using IVF
* Total 2 reports the minimum number of respondents who provided one or more inconsistencies

Table 7.2 summarises the main results. Out of the 5,739 respondents, 3,040 (53.0%) stated to be willing to use IVF if it were free. Therefore, 47% of respondents attributed no value to IVF and would not use it even in the case it were free. 2,482 (43.2%) of the respondents were willing to pay Italian Liras 1 million (€ 516), this was 558 less than those who were willing to use it if it were free. For higher prices the number of respondents willing to pay for IVF drops further; for the highest value (Italian Liras 50 million that it is about € 25,800) the number of "users" is 500, about 8.7% of the total number of respondents. Overall, as predicted using standard economic theory, the higher the bid the fewer the number of respondents willing to pay.

Table 7.2. Willingness to pay for using IVF in case of infertility (yes/no answers to 6 bids).

	Willing to Pay	No Willing to Pay
Bid	# (%)	# (%)
It. Liras 0	3.040 (53.0)	2.699 (47.0)
It. Liras 1 million (€ 516)	2.482 (43.2)	3.257 (56.8)
It. Liras 5 million (€ 2,582)	1.895 (33.0)	3.844 (67.0)
It. Liras 10 million (€ 5,165)	1.309 (22.8)	4.430 (77.2)
It. Liras 20 million (€ 10,329)	879 (15.3)	4.860 (84.7)
It. Liras 50 million (€ 25,823)	500 (8.7)	5.239 (91.3)

On the basis of the data generated by these answers we graphed the basis of a demand schedule of IVF for personal use (figure 7.1). As we only observed 6 points of the curve, we assumed that demand is linear between each consecutive pairs of points and we created two scenarios: one with the maximum WTP set at Italian Lira 50 million (€ 25,823) and the other assuming that the number of respondents willing to pay linearly declines at a constant rate. Visual representations of the demand schedule for personal use of IVF for the two scenarios are presented in figure 7.2. and figure 7.3.

Figure 7.1. Number of respondents willing to pay for IVF at different bids

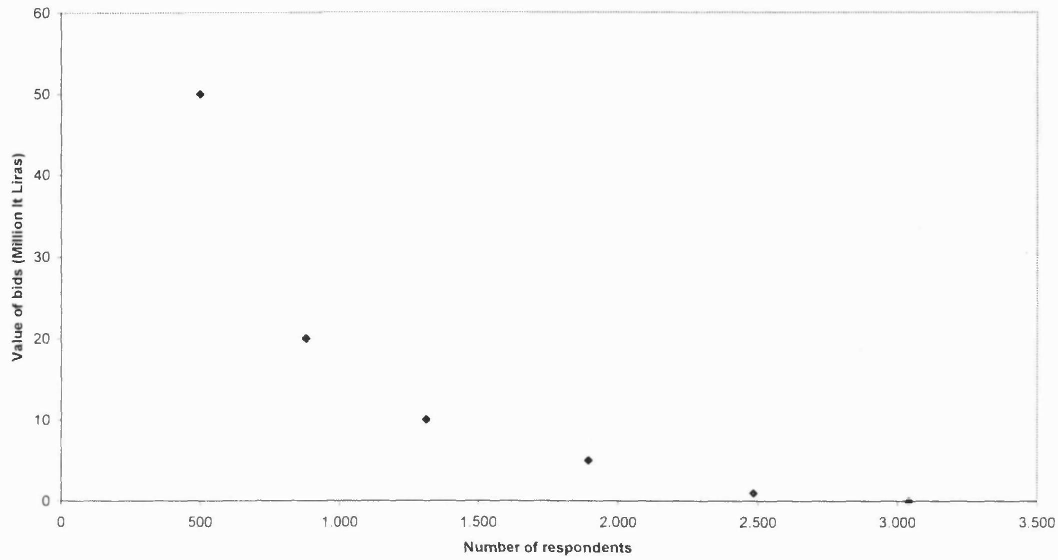


Figure 7.2. Demand schedule for IVF
(assuming no respondent is willing to pay more than Italian 50 Million - Euro 25,823)

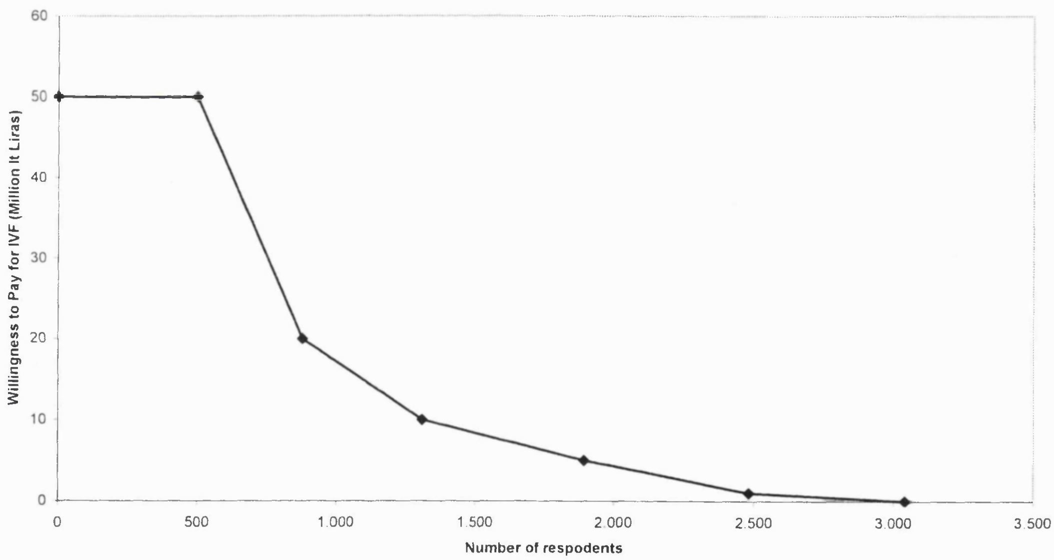
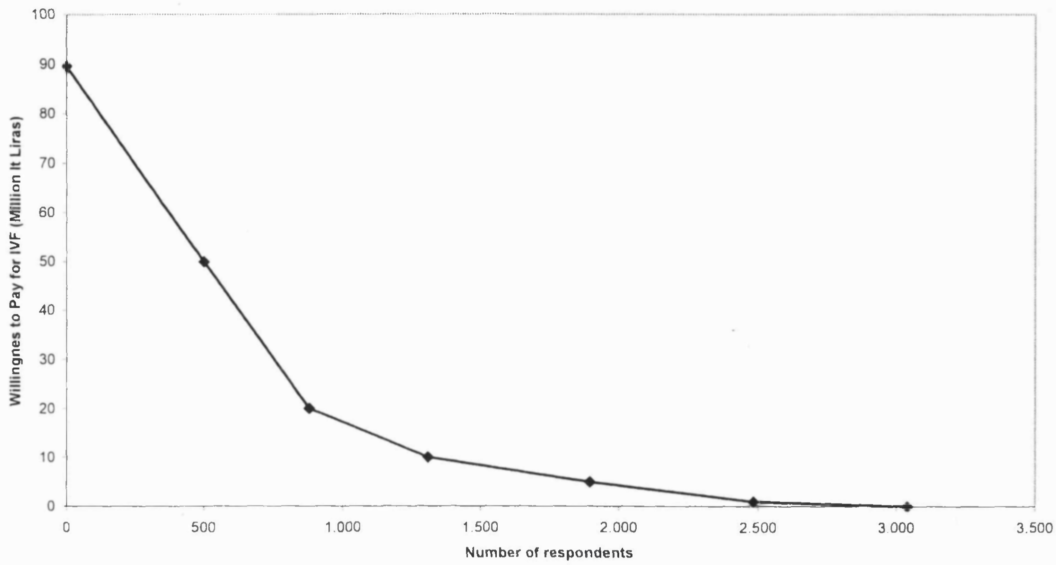


Figure 7.3. Demand schedule for IVF
(assuming linear decrease in the number of respondents)



Total WTP is the area beneath the curve. If it is assumed that It. Liras 50 Million (€ 25,823) is the maximum WTP value, the aggregate WTP of the sample is It. Liras 52,980 Million (€ 27.4 Million), that is It. Liras 9.23 Million (€ 4,767) per respondent and It. Liras 17.43 Million (€ 8,999) per user (respondent willing to use IVF). Under the other assumption (linear decrease of the number of persons willing to pay more that It Liras 50 million) total WTP is It. Liras 61,025 million (€ 31.5 million) and the WTP per respondents and per user are It. Liras 10.6 million (€ 5,474) and It. Liras 20.1 Million (€10,381), respectively.

In summary, this part of the survey shows that the mean willingness to pay for IVF of those respondents who would use IVF in case of need would range from € 9.000 to € 10.381. This is the value attributed to an IVF package providing up to 3 IVF cycles with an overall probability of success (life birth) of 30%.

7.3. Willingness to pay for a publicly funded programme providing IVF to infertile couples

We first administered the TIOLI questions by splitting the total sample into four sets. Each sub-sample consisted of about 1,400 individuals (minimum 1,384 and maximum 1,490 (table 7.3). Respondents who chose the option "Do not know" to the TIOLI question were about 15% of the sample. In the "It. Liras 5,000" (€ 2.6) sub-sample, 29% of respondents were *definitely* willing to pay for the programme. For the same amount another 25.2% of respondents were *probably* willing to pay. Overall "yes" answers in this sub-sample accounted for 54.2% of respondents. For the highest bid, that is It. Liras 50.000 (€ 25.8), the proportion of respondents willing to pay was 41.1% (26.3% *probably* and 14.8% *definitely*). The proportion of "Yes, definitely" answers is strictly decreasing and the proportion of "No, definitely not" answers is strictly increasing. However, both *probably* answers are not monotone and as a consequence both overall "yes" and "no" are not monotone as well. The proportion of respondents willing to pay It. Liras 10,000 (€ 5.2) is 55.2%, a little higher than 54.2% that is the proportion willing to pay It. Liras 5,000 (€ 2.6).

Table 7.3. Willingness-to-pay for a publicly funded programme providing IVF to infertile couples: referendum format (each bid presented to one of the four sub-samples)

	Bid in Italian Liras (€)									
	It Liras 5,000 (€ 2.6)		It. Liras 10,000 (€ 5.2)		It. Liras 20,000 (€ 10.3)		It. Liras 50,000 (€ 25.8)		All bids	
	#	%	#	%	#	%	#	%	#	%
Do not know	213	14,30%	202	14,03%	275	19,87%	211	14,81%	901	15,70%
Yes, probably	376	25,23%	418	29,03%	407	29,41%	374	26,25%	1575	27,44%
Yes, definitely	432	28,99%	377	26,18%	254	18,35%	211	14,81%	1274	22,20%
No, probably not	213	14,30%	189	13,13%	199	14,38%	303	21,26%	904	15,75%
No, definitely not	256	17,18%	254	17,64%	249	17,99%	326	22,88%	1085	
Yes (probably or definitely)	808	54,23%	795	55,21%	661	47,76%	585	41,05%	2849	49,64%
No (probably or definitely)	469	31,48%	443	30,76%	448	32,37%	629	44,14%	1989	34,66%
Total	1490		1440		1384		1425		5739	

"Do not know" excluded	Bid in Italian Liras (€)									
	It Liras 5,000 (€ 2.6)		It. Liras 10,000 (€ 5.2)		It. Liras 20,000 (€ 10.3)		It. Liras 50,000 (€ 25.8)		All bids	
	#	%	#	%	#	%	#	%	#	%
Yes, probably	376	29,44%	418	33,76%	407	36,70%	374	30,81%	1575	32,55%
Yes, definitely	432	33,83%	377	30,45%	254	22,90%	211	17,38%	1274	26,33%
No, probably not	213	16,68%	189	15,27%	199	17,94%	303	24,96%	904	18,69%
No, definitely not	256	20,05%	254	20,52%	249	22,45%	326	26,85%	1085	22,43%
Yes (probably or definitely)	808	63,27%	795	64,22%	661	59,60%	585	48,19%	2849	58,89%
No (probably or definitely)	469	36,73%	443	35,78%	448	40,40%	629	51,81%	1989	41,11%
Total	1277		1238		1109		1214		4838	

This unexpected pattern is due to very few cases and may be attributable to chance. Nevertheless, it is inconsistent with standard economic theory that assumes that the higher the price the fewer the number of consumers. The results of this part of the survey do not fully support the validity of the TIOLI format. In addition, while the highest bid (It. Liras 50,000) is 10 times the lowest one (It. Liras 5.000) the proportion of yes answers (*probably* and *definitely*) drops only by an absolute 13% (from 54.2% to 41.1%). Respondents do not appear to be very sensitive to the value of the bid so that it may be argued that the TIOLI questions produce inaccurate estimates.

Mean WTP values calculated according to different hypotheses are presented later in this chapter. In the meanwhile it can be noted that if this data derived from a real referendum and a strict majority rule applied (absolute majority of voters), a tax of It. Liras 10,000 (€ 5.2) to fund IVF would have been approved. If a relative majority sufficed to approve the proposal (a less stringent rule), a tax of It. Liras 20,000 (€ 10.3) would have been approved.

The modified payment card (MPC) questions were the same for all respondents and followed the TIOLI questions. Respondents were presented 8 values, from It. Liras 2,000 (€ 1) to It. Liras 200,000 (€ 103.3). Values were presented randomly to each respondent. For each value respondents were required to choose between yes and no without the "*do not know*" option and the possibility to graduate the answers by the use of *probably* and *definitely.*

Out of the 5,739 returned questionnaires 581 (10.1%) were deemed inconsistent as they reported "yes" to bids that were higher of bids for which they reported "no". This relatively high number of inconsistent answers is probably due to the random presentation of bids. As respondents faced 8 unordered values unintentional mistakes are likely and may reflect uncertainties and difficulties of the cognitive processes. Eight-nine questionnaires (1.5% of the total) twice violated the positive association between price and willingness to pay (that is a sequence no-yes-no-yes to increasing values of willingness to pay). This value is significantly higher than those observed in the MPC questions for private use of IVF. This may due to the higher number of values offered (8 compared to 6 in the part on personal use of IVF) or by the fact that the part of the questionnaire on WTP for the publicly funded programme was administered at the end, when respondents were probably more tired and less concentrated.

All 5,739 answers, including those inconsistent were kept in the data set to be analysed in the next sections. While there is uncertainty about how to treat apparently invalid data, we preferred not to exclude any observation in order to fully reflect what respondents answered. Almost 56% of respondents stated to be willing to pay at least It. Lira 2,000 (Euro 1) (table 7.4). This percentage drops for higher values: it is about 53% for It Lira 10,000 (€ 5.2), 32% for It. Lira 50,000 (€ 25.8) and 10.2% for It. Lira 200.000 (€ 103.3).

Table 7.4. Willingness-to-pay for a publicly funded programme providing IVF to infertile couples: modified payment card format (all bids presented to the entire sample)

	Willingness-to-Pay			
	Yes		Not	
Bid values	#	%	#	%
It. Liras 2,000 (€ 1.0)	3,192	55,62%	2,547	44,38%
It. Liras 5,000 (€ 2.6)	3,168	55,20%	2,571	44,80%
It. Liras 10,000 (€ 5.2)	3,043	53,02%	2,696	46,98%
It. Liras 20,000 (€ 10.3)	2,632	45,86%	3,107	54,14%
It. Liras 50,000 (€ 25.8)	1,845	32,15%	3,894	67,85%
It. Liras 80,000 (€ 41.3)	1,291	22,50%	4,448	77,50%
It. Liras 100,000 (€ 51.6)	1,127	19,64%	4,612	80,36%
It. Liras 200,000 (€ 103.3)	590	10,28%	5,149	89,72%

For the maximum value offered (€ 103.3) 590 respondents, that is about 10.3% of the sample, stated to be willing to pay. Such a large number of respondents for the highest value offered is unexpected and suggests that the sample may be censored. It is likely that a significant number of respondents would have been willing to pay more that € 103; consequently, to set this value as the upper WTP limit may underestimate the maximum willingness to pay of many respondents.

On the other side 44.3% of the sample stated they were not willing to pay even € 1, the minimum value offered. We attributed 0 WTP to these individuals. However, it is plausible that some of these patients may have a negative willingness to pay for a publicly funded IVF programme as ARTs present controversial ethical issues. Therefore, it looks appropriate to assume that there is a spike at zero or that observations are left-censored because the latent dependent variable may assume negative values.

7.4. WTP estimates elicited from the referendum format

As detailed in the previous chapter, we assumed that the probability that an individual will say "yes" to the WTP question follows a logistic distribution and we estimated mean WTP as the area under the probability function. This area shows the fraction of the sample who would consume the good at each price (bid) level. Therefore mean willingness to pay can be estimated by integrating the probability function.

$$E(WTP) = \int_L^U (1 + e^{-a+b*bid})^{-1} db - \int_L^0 1 - (1 + e^{-a+b*bid})^{-1} db$$

Where a and b are the coefficients of the estimated regression logit equation when only the bid amount (bid) is included, U is the upper limit of the integration and L the lower limit of the integration. Table 7.5 presents the coefficients and the limits used for the integrations used in the different scenarios.

Table 7.5. Logit regression results of the take-it-or-leave-it data and estimates of mean willingness to pay.

				A	B	C	D
Lower Limit				€ 0	€ 0	€ -25.8	€ -25.8
Upper Limit				€ 25.8	€103.3	€ 25.8	€ 103.3
Mod.	No/Yes	Const.	B coeff.	Mean willingness to pay			
1	1989/2849	0.67409	-0.01475	€ 19.04	€ 89.77	€ 8.06	€ 78.80
2	2890/2849	0.25089	-0.01264	€ 16.96	€ 82.22	€ 3.64	€ 68.89
3	1989/1274	0.13940	-0.00218	€ 14.16	€ 60.73	€ 0.99	€ 47.56
4	4465/1274	-0.8818	-0.01916	€ 10.39	€ 71.54	€ -10.06	€ 51.09

LL = lower limit of the integration; UL = upper limit of the integration; It L. 50,000 = € 25.8, It L. 200,000 = € 103.3.

The four rules for reducing to a binary format the answers collected in the questionnaire produce different mean WTP (see section 5.7). If the limits of the integration are those corresponding to the lowest and the highest bids offered to respondents (0 and It. Liras 50,000), mean WTP is about € 19 for model 1 ("do not know" treated as missing, "yes, definitely" and "yes, positively" treated as "yes" and "no, definitely not" and "no, positively not" treated as "no") and is € 10.40 for the most conservative scenario (model 4 in which only "yes, definitely" is treated as "yes" and all the other answers are treated as "no"). As expected, for larger integration intervals mean WTP estimates are larger; if the probability function is integrated between negative Italian Liras 50,000 (€ 25.80) and Italian Liras 200,000 model 1 and model 4 produce mean WTP equal to € 78.80 and to € 51.10, respectively.

The assumptions about the integration limits are crucial. In model 1 results vary more than tenfold, from € 8.10 to € 89.80; in model 4, the most conservative, they vary from a negative value of € 10.10 to a positive value € 71.50. It is important to investigate why the integration limit assumptions are so important. Almost 1/3 of respondents refused to pay the lowest value offered. In a previous section of the questionnaire it was revealed that 47% of respondents were not willing to pay anything for IVF in case of infertility. These respondents would not have used IVF even if it were completely free of charge. Although altruistic feelings may cause some of these respondents to be willing to pay a positive amount even if they excluded their personal use (see below), it looks very likely that many respondents who are not willing to pay € 2.60 (the lowest value offered) have no willingness to pay for the publicly funded IVF programme. Moreover, it is also likely that some people do not derive any utility from the programme. These arguments suggest setting the lower integration limit at 0.

However, as mentioned in the previous chapter, the case of ARTs is so peculiar that some individuals may experience disutility as a result of the publicly funded programme providing IVF. Therefore, for this specific question negative WTP is possible and thus setting a negative lower integration limit may be plausible. However, it should be made clear that the setting of such a value would be arbitrary as no evidence was collected on the magnitude of the disutility of the IVF programme for any respondent.

Table 7.5 clearly shows that the assumptions about the lower integration limits have a great impact on results. In model 1 changing the lower integration limit from 0 to negative It. Liras 50,000 (€ -25.80) implies a reduction in mean WTP from € 19 to € 8.10 if the upper bound is It. Liras 50,000 (€ 25.80) and a reduction in mean WTP from € 89.80 to € 78.80 if the upper bound is Liras 200,000 (€ 103.30). In model 4, the very conservative one, mean WTP for the IVF programme is negative if the upper value is It. Liras 50,000 (€ 25.80) and the lower value is It. Liras 50,000 (€ -25.80) (see column c in table 7.5).

The assumption about the upper integration limit is crucial as well. Out of the 1,425 respondents to the It. Liras 50,000 (€ 25.80) group, 374 and 211 were probably and definitely willing to pay, respectively. Almost 15% of the sub-sample stated to be sure to be willing to pay the highest amount offered. It is thus very likely that some of them would have been probably (and even definitely) willing to pay higher amounts. This justifies the hypothesis of the upper bound set at It. Liras 200,000 (€ 103.30). Such a hypothesis has a great impact on results. In model 1 raising the upper limit of the integration raised mean WTP from € 19 to € 79.80 if the lower integration limit is 0, and from € 8.10 to € 78.60 if the limit is It. Liras -50,000.

The referendum format questions used to elicit willingness to pay required a few assumptions about how to treat "do not know" answers, how to code the distinction between "definitely" and "probably" answers and about the minimum and the maximum willingness-to-pay admissible. Our results show that this last set of assumptions, that is the definition of the upper and the lower limits of the distribution are more relevant than those related to coding. However, it should be clear that the high sensitivity of results to the assumption of minimum and maximum WTP derives from the very limited number of bids used in the present survey (only four in this part of the questionnaire) and the limited sensitivity of respondents to the range of value offered. Excluding "do not know", the percentage of respondents willing to pay varies from 63.3% for the lowest value offered (€ 2.60) to 48.2% for the highest one (€ 25.80). Consequently, the survey was not very informative. It provided limited information on the 36.7% of respondents whose WTP is less than € 2.60 and on the 48.2% of respondents whose WTP is likely to be higher than € 25.80.

Keeping in mind these limitations, results presented here can be also interpreted according to a policy perspective. Depending of the referendum rules, the survey reveals different implications. If the proposal had to be approved by the absolute majority of voters, a tax of € 5.20 to fund IVF would have been approved while taxes of higher values would have been rejected. Results are more favourable to the IVF programme if a less stringent majority rule is used. If the proposal had to be approved by the majority of those expressing a yes/no vote (thus excluding "do not know" from computation) a clear majority of 58.9% of votes would have been in favour of raising a tax of € 10.30 to fund IVF. However, the referendum logic neglects the intensity of preferences. Instead, the cost-benefit analysis needs benefit measures reflecting the magnitude of utility generated by public choices. According to a cost-benefit perspective it is thus more relevant to calculate mean WTP than the WTP of the median voter.

We think that model 1 under the hypothesis a) is the most plausible. Model 1 uses both the "probably" and the "definitely" answers respecting the sign (yes and no) and integrate the probability function between two limits that are directly observed. The use of these limits may inflate mean WTP as the lower limit excludes the possibility of negative WTP but, at the same time, reduces mean WTP as many respondents may be willing to pay amounts that are much higher than those offered in the survey. According to these assumptions, mean WTP is about € 19. If different assumptions are made about coding of "do not know" and "yes, probably" answers, but the integration limits are not varied, mean WTP ranges from € 10.40 to € 17. In section 6.13 we compare these benefits to estimates of cost for producing IVF services.

7.5. Spike models

A major characteristic of our dataset is that a large number of respondents answered "no" to any proposed value. From this we can draw the conclusion that most of these individuals were not in the market in the sense that attributed zero value to the public programme. If this is so, it may be inappropriate to assume that the probability that a respondent is willing to pay a certain amount is given by a logistic probability function. It is likely that a more articulated assumption about the probability distribution is needed. One option, following Kristrom (1997), consists of assuming that the panel is split in two parts, one for whom the programme provides benefits (even if marginal) and the other for whom the programme does not produce any benefit. As presented earlier, for such a model we have to estimate parameters from a composite function: In essence, the model first describes whether the respondent is willing to pay any amount for the programme and then estimates the parameters of the probability distribution function of willing to pay as a function of the proposed bid. Table 6 presents the results of these analyses.

Table 7.6. Parametric estimates of mean WTP: the effects of zero WTP

		Logit Lower limit -∞ Upper limit + ∞			Spike Logit Lower limit 0 Upper limit + ∞		
	No/Yes	Const.	B coeff.	Mean Willingness to Pay (€)	Const.	B coeff.	Mean Willingness to Pay (€)
1	1989/2849	0,67409	-0,01475	23.60	0.09668	0.14752	26.00
2	2890/2849	0,25088	-0,01264	10.25	0.08799	0.12645	30.34
3	1989/1274	0,01394	-0,02284	0.31	0.65798	0.02284	16.80
4	4465/1274	0,01394	-0,01916	-23,77	0.96697	0.01915	20.10

Estimates on WTP through a parametric logit model are regularly below those having a spike at zero. This is because the full parametric model requires the integration of the overall function in the real domain, including negative values. Therefore, zero values are understood by the model as truncated observations on negative values. Consequently, the entire distribution of WTP, compared to the spike models, is shifted towards left and consequently mean WTPs are lower. The spike model estimates that about 27% of observations are true zero WTP (this is approximately the percentage of respondents that were against any funding of a public programme providing IVF) and provide WTP estimates between €16 and € 31.

It is important to note that the two main methodological issues are crucial here: the form of the probability distribution function and the limits of integration. We only proposed positive values; however, we have used the answer to another question to detect respondents who have true WTP for the programme. This data can be used according to a large variety of assumptions about the distribution of probability (here the distribution of interest is the probability of providing yes to a specific bid). If we use a logistic distribution we implicitly recognize that the probability distribution function may be positive (with probability greater than zero) also for negative WTP values. Instead, by using the spike model we assume that the negative realm is excluded and that a relative large amount of respondents have zero WTP.

The other relevant issue is the interval of integration. If we use a parametric approach, this interval should be drectly identified by the probability distribution function; otherwise the mathematical result has unclear statistical interpretation (Hanemann and Kanninen; 1996). Consequently, in the parametric model without any spike the integration linterval covers negative values and thus mean WTP results lower. What drives down mean WTP in the model without spike is the number of observations at zero. Results of assumption 4 (table 7.6) presents the largest difference between the model with and without the spike because the large number of zero generate a high number of negative WTP value in the former and zero WTP in the latter. The higher the number of observations at zero (as in assumption 4) the larger the fraction of observations that contribute differently to the shape of the distribution and thus to the mean value (but not to the median values).

7.6. WTP estimates elicited from the modified payment card format

Table 7.6 presents data and calculations to obtain total sample and mean WTP from the modified payment card data. Respondents are attributed to the value that they reported as the maximum WTP. For each value it is also presented the value used to calculate WTP (the mid-point between the amount at which the respondent said "yes" and the amount at which he/she said "no").

Table 7.7. Total and mean WTP for a public programme funding IVF to infertile couples of the total sample (5,739 respondents).

Amount offered	Mid-range value (*)	Number of respondents willing to pay	Total Willingness to pay
It. Liras 0	0	2,057	€ 0
It. Liras 2.000 (€ 1)	€ 1.80	109	€ 196
It. Liras 5.000 (€ 2.60)	€ 3.90	251	€ 979
It. Liras 10.000 (€ 5.20)	€ 7.80	474	€ 3,697
It. Liras 20.000 (€ 10.30)	€ 18.10	869	€ 15,728
It. Liras 50.000 (€ 25.80)	€ 33.60	561	€ 18,849
It. Liras 80.000 (€ 41.30)	€ 46.50	251	€ 11,420
It. Liras 100.000 (€ 51.60)	€ 77.50	577	€ 44,717
It. Liras 200.000 (€ 103.30)	€ 103.30	590	€ 60,947
Total WTP (€)			€156,536
Mean WTP per respondent (€)			€ 27.28

(*) Maximum WTP = 0 for respondents not willing to pay It. Liras 2,000 and maximum WTP = 200,000 for those willing to pay It. Liras 200,000.

The sample has a total WTP of about € 156,536. The highest valued offered (€ 103.30) contributes to total WTP for almost 40% of the total. As in the referendum question, the high number of respondents willing to pay the highest amount offered has a decisive impact on the WTP estimate and raises the issue of the WTP amount to be attributed to those willing to pay the maximum value proposed. A consistent minority (about 10% of the sample) is probably willing to pay amounts higher than those presented in the questions and strongly influences the sample mean value.

Mean WTP, estimated as the ratio between total WTP and total number of respondents (5,739), is € 27.28. This means that, on average, respondents of this survey are available to pay € 27.28 to fund a national programme providing free IVF to infertile couples.

The survey over represent some population groups (e.g. the well educated) and under represent other groups (e.g. the elderly). In order to have results more representative of the Italian population each respondent was assigned a weight to adjust his/her relative importance in the sample. Table 7.8 reports total and mean WTP of the weighted sample. Also, it reports the average weight of the respondents for each of the 8 values offered. The weighting procedure has a limited impact on results: total WTP is € 152,580 (it is € 156,536 if data are not weighted) and mean WTP is € 26.59, about € 0.70 less than the value obtained without weighting.

Table 7.8. Total and mean WTP for a public programme funding IVF to infertile couples of a sample representative of the Italian population in respect of age, gender, education, employment status, geographical area and municipality size of respondent' residence.

Amount offered	Mid-range value (*)	Number of respondents willing to pay	Average weight	Total willing to pay
It. Liras 0	0	2,057	1.174901	€ 0
It. Liras 2.000 (€ 1)	€ 1.80	109	1,224666	€ 240
It. Liras 5.000 (€ 2.60)	€ 3.90	251	0.985312	€ 964
It. Liras 10.000 (€ 5.20)	€ 7.80	474	0.896577	€ 3,315
It. Liras 20.000 (€ 10.30)	€ 18.10	869	0.919385	€ 14,460
It. Liras 50.000 (€ 25.80)	€ 33.60	561	1.282328	€ 24,171
It. Liras 80.000 (€ 41.30)	€ 46.50	251	1.083753	€ 12,377
It. Liras 100.000 (€ 51.60)	€ 77.50	577	0.896416	€ 40,085
It. Liras 200.000 (€ 103.30)	€ 103.30	590	0.934676	€ 56,966
Total WTP (€)				€152,580
Mean WTP per respondent (€)				€ 26.59

(*) Maximum WTP = 0 for respondents not willing to pay It. Liras 2,000 and maximum WTP = 200,000 for those willing to pay It. Liras 200,000.

These values are substantially lower than the estimates obtained from the referendum questions (about € 70-80) if the upper integration limit is set at Italian Lire 200,000 (€ 103.30). However, if the integration limit is set at the highest bid proposed in the TIOLI questions (Italian Liras 50,000 that corresponds to € 25.80), the estimate derived from the MPC questions are substantially higher (€ 27.28 compared to € 10-20).

7.7. Egoistic and altruistic willingness to pay

We estimated that, in a hypothetical case of infertility, respondents willing to try IVF (potential users of IVF) have a mean willingness to pay of about € 9-10,000 for 3 cycles of IVF. Then, from other questions of the survey we estimated that mean WTP for a national programme providing free IVF to infertile couples is likely to be in a € 10-30 range.

The first estimate (€ 9-10,000) refers to the use of IVF and thus reflects WTP for the personal use of the intervention. This estimate should only include what the person is willing to pay in the case he/she needs the intervention. Therefore, this value does not include any option value and any altruistic component. It does not include any option value because respondents are presented an ex-post situation (*..imagine that you are infertile...*) and it does not include any altruistic component because the question referred to WTP for the personal use of the service with no reference to any collective scheme of funding.

The second WTP value, elicited through the take-it-or-leave-it and the modified payment card formats, refers to a publicly funded programme providing IVF to infertile Italian couples. This WTP value should include WTP for personal use, WTP for the option of having WTP in case of need and WTP for making it freely available to other members of the community. Therefore, this WTP value is expected to capture the altruistic component of IVF.

Unfortunately, a direct comparison between the two estimates cannot be performed. Respondents were not given their probability of being infertile and did not have enough information to make an estimate of expected probability of using IVF. The provision of this information would have made the questionnaire too complicated and to cognitively difficult. In addition, about 40% of respondents were not of fertility age and even those who were of fertility age had different probability of being infertile due to their specific age.

Despite the missing link between the two WTP estimates it is possible to further investigate the data in order to understand the magnitude of the altruistic component of WTP. As one question asked whether the respondents would have used IVF in case of infertility, it is possible to provide separate WTP estimates for potential user and non users of IVF (table 7.9).

Table 7.9. Personal use of IVF and willingness-to-pay for a national programme funding IVF to infertile couples

	"Imagine you are infertile and you desire a baby. Would you try IVF?"						
	Do not know	No, definitely not	No, probably not	Yes, probably yes	Yes, definitely yes	Not	Yes
Amount offered	Number of respondents willing to pay						
It. L 0	552	733	488	211	73	1221	284
It. L 2.000 (€ 1)	23	18	21	37	10	39	47
It. L 5.000 (€ 2.60)	55	41	57	60	38	98	98
It. L 10.000 (€ 5.20)	106	36	78	191	63	114	254
It. L 20.000 (€ 10.30)	134	69	118	419	129	187	548
It. L 50.000 (€ 25.80)	76	26	82	250	127	108	377
It. L 80.000 (€ 41.30)	38	18	26	118	51	44	169
It. L 100.000 (€ 51.60)	69	25	66	269	148	91	417
It. L 200.000 (€ 103.30)	66	32	44	250	198	76	448
Total respondents	1.119	998	980	1.805	837	1.978	2.642
Mean WTP per respondent	17.83	8.68	16.94	38.68	49.59	12.77	42.13
% of respondents having positive WTP	50.7 %	26.6%	50.2%	88.3%	91.3%	65.5%	83.8%

Using the non-parametric approach to estimate the WTP from the MPC data, the mean WTP for the publicly funded IVF programme is much larger among users of IVF (€ 42.13) than among non users (€ 12.77). Moreover, the mean WTP is higher among those being *definitely* in favour of using IVF (€ 49.59) than among those being *probably* in favour (€ 38.68); on the contrary, respondents that were *definitely* against the use of IVF have a lower WTP (€ 8.68) than those who were *probably* against its use (€ 16.94).

These results were expected because respondents who would not use IVF, and thus have 0 WTP for personal use of IVF, should state a lower WTP for a publicly funded programme. This is because for the respondents who are against personal use of IVF the entire value of the programme is generated by altruism. Almost two-thirds of the respondents who were not in favour of using IVF would have paid something for a public programme providing IVF to infertile couples. Mean WTP of these respondents (those who would not use IVF), provides an estimate of the altruistic component of total WTP. The approximate mean WTP of € 12.77 for these respondents can be used as a proxy of the altruistic component of WTP for the programme (about 50% of mean WTP of the entire sample).

The fact that almost two-thirds of the sample with no intention to use IVF would have paid something for IVF and that their mean WTP is substantial provide an important indication that altruism is a relevant determinant of willingness-to-pay for a public programme in the area of assisted reproduction.

7.8. Validity considerations

As respondents were split into four groups according to the amount offered in the TIOLI question, we could check if different amounts have an impact on the answer to the MPC question (see previous chapter). Estimates of mean WTP ranges from € 25.50 € to € 29.40 and tend to be positively associated to the value offered in the TIOLI question (table 7.8). The higher the value offered, the higher the mean WTP elicited from modified payment card data. In a linear regression where the dependent variable is mean WTP, the coefficient of the value offered in the TIOLI question is significantly positive (t=2.50, p=0.012). It can be stated that the TIOLI question had an impact on the WTP elicited from the MPC questions. The magnitude of the impact is significant but not large: mean WTP for the respondents attributed to the lowest value offered (about € 2.50) is about € 4 less than the mean WTP for the respondents attributed to the highest value (€ 25.80). This absolute variation is about 15% of the total sample mean and suggests the existence of an anchoring effect. Once respondents are given an amount to decide whether they are willing to pay, they are influenced by that amount when answering to additional questions.

Table 7.10. Mean WTP from the modified payment card (MPC) data: comparison of the four sub groups used in the take-it-or-leave-it question.

Sub-group	Mean WTP (including do not know to the TIOLI question) (a)	Mean WTP (excluding do not know to the TIOLI question) (b)
It. Liras 5,000 (€ 2.60)	It. Liras 46.470 (€ 24.00)	It. Liras 55.310 (€ 28.60)
It. Liras 10,000 (€ 5.20)	It. Liras 51.311 (€ 26.50)	It. Liras 60.772 (€ 31.40)
It. Liras 20,000 (€ 10.60)	It. Liras 51.698 (€ 26.70)	It. Liras 61.551 (€ 31.80)
It. Liras 50,000 (€ 25.80)	It. Liras 54.021 (€ 27.90)	It. Liras 61.452 (€ 31.70)
All sub groups	It. Liras 52.821 (€ 27.30)	It. Liras 59.696 (€ 30.80)

The last column of table 7.9 reports the same analysis for a smaller dataset: all "do not know" answers to the TIOLI questions are excluded. Mean WTP is higher for respondents who were attributed to higher bids even if "do not know" respondents are excluded, although the magnitude of the effect appears slightly smaller.

The exclusion of undecided respondents is associated with higher mean WTP (€ 30.80 versus € 27.30). In other words, if respondents who did not know whether they were willing to pay certain amounts for a publicly funded IVF programme in the TIOLI question are not included to calculate mean WTP from MPC data, the estimate is higher. This suggests that the inclusion of the "do not know" option is not neutral and tend to capture more "yes" respondents than "no" respondents. Therefore, including "do not know" makes the WTP elicitation method more conservative.

7.9. A comparison between the two elicitation methods.

We compared the TIOLI and the MPC elicitation methods in various ways. According to the standard procedure presented above mean WTP from MPC data is estimated to be € 27.28. As the MPC format did not allow the possibility of negative WTP, a meaningful comparison of mean WTP from the two formats must exclude negative integration limits for the analysis of TIOLI data. If the lower integration limit is assumed to be 0, mean WTP calculated from TIOLI data ranges from € 10.40 to € 89.80 (table 7.5). As reported above, mean WTP estimates derived from TIOLI data are sensitive to the integration limits and coding of "do not know", "probably" and "definitely" answers. If the upper integration limit is Italian Liras 200,000 (€ 103.30), the highest value offered in the MPC approach, mean WTP calculated from the TIOLI data varies from € 60.7 to € 80.8. These values are substantially higher than the mean WTP derived from MPC data (€ 27.30).

While this comparison assumes that the upper integration limit to calculate mean WTP is equal to the highest value offered in the MPC, a similar comparison can be conducted by calculating mean WTP from the MPC data including only answers to the value that were also used in the TIOLI question. This implies calculating mean WTP assuming that yes/no answers are available only for 4 values offered to respondents (It Liras 5,000, 10,000, 20,000 and 50,000). According to this assumption, mean WTP for the MPC data is €10.70, a value that is smaller than those derived from the TIOLI data. Only if a very conservative assumption is made, which consists in assuming that only "definite yes" reveal that respondents are willing to pay and that all the other answers (including "yes, probably") reveal that they are not, mean WTP derived from TIOLI data is similar to mean WTP derived from MPC data. However, if model 1, 2 or 3 are assumed (table 7.5), it clearly appears that TIOLI data produce larger estimates of mean WTP.

Finally, we simulated TIOLI data from those elicited with the MPC method and made a comparison between actual and simulated data (table 7.11). Simulated TIOLI data are very similar to those observed. If “do not know” answers are excluded, as they cannot be simulated, the percentage of total “yes” answers is very similar between observed and simulated data (58.8% and 58.7%, respectively). In the simulated data the percentage of “yes” to the highest bid (Italian Liras 50,000 - € 25.88) is slightly higher; on the contrary, the percentages of observed “yes” for the lower bids are higher than those calculated from the simulated data.

Table 7.11. Data collected from the TIOLI question and data generated by simulating TIOLI data from the answers to the modified payment card questions.

Bid	Observed TIOLI data		Simulated TIOLI data (including "Do not know" respondents to TIOLI question)		Simulated TIOLI data (excluding "Do not know" respondents to the TIOLI question)	
	Willing to Pay	No Willing to Pay	Willing to Pay	No Willing to Pay	Willing to Pay	No Willing to Pay
	# (%)	# (%)	# (%)	# (%)	# (%)	# (%)
It. Liras 5,000 (€ 2.60)	808 (63.27)	469 (36.76)	893 (59.93)	597 (40.07)	830 (65.00)	447 (35.0)
It. Liras 10,000 (€ 5.20)	795 (64.22)	443 (35.78)	873 (60.63)	567 (39.38)	812 (65.59)	430 (65.59)
It. Liras 20,000 (€ 10.60)	661 (59.60)	448 (40.40)	721 (52.10)	663 (47.90)	657 (59.24)	452 (40.76)
It. Liras 50,000 (€ 25.80)	585 (48.19)	629 (51.81)	586 (41.12)	839 (58.88)	541 (44.56)	673 (55.44)
Total	2,849 (58.89)	1,989 (41.11)	3,073 (53.55)	2,666 (46.45)	2,686 (58.70)	2,152 (41.30)
Logit estimates (constant and bid coefficient)	a = 0.67409 b = -0.01475		a = 0.52002 b = -0.01790		a = 0.77392 b = -0.01973	
Mean WTP	It. Liras 36,866 (€ 19.04)		It Liras 38,821 (€ 20.04)		It Liras 36,083 (€ 19.00)	

For both the simulations the same assumption used for the original TIOLI data were made concerning the distribution of the probability that an individual will say "yes" to the WTP question. It was assumed that it follows a logistic distribution:

$$P(yes) = (1 + e^{-a+b*bid})^{-1}$$

As shown above, mean WTP can be calculated from the probability function by integration. The comparison of the original and simulated data is based on the assumptions that the minimum WTP is 0 and that the maximum is It. Liras 50,000 (€ 51.60). Therefore:

$$E(WTP) = \int_0^{50} (1 + e -a+b*bid)^{-1}db$$

Results derived from observed and simulated data are strikingly similar. Simulated TIOLI data generate WTP estimates ranging from €19 to € 20. Based on the same assumptions mean WTP from the observed data is € 19.04. If data collected from the modified payment card approach are transformed in TIOLI data, the distribution of "yes" in the simulated data is similar to the distribution of original TIOLI data. WTP estimates are also very similar. Two main reasons may drive these results. First, estimates from MPC data may be higher than those derived from TIOLI data because of the assumptions concerning the probability function needed to calculate mean WTP. The second reason why actual and simulated data are similar may derive from the sequence of the questions. Respondents answered the TIOLI question before the MPC one. Thus it is plausible that they have anchored their answers to the second question to the first one. In other words, simulated and actual TIOLI data are similar because respondents wanted to be consistent across the two questions.

7.10. Cost estimates

In this section we present the results of cost analysis. The justification of the general approach adopted is presented in Chapter 4 and details about the source of data and the methods used are described in Chapter 5.

Before presenting the results of the cost analysis it is important to underline the general approach which was followed. We used a full costing methodology as we estimated the cost of IVF cycle as the sum of direct and indirect costs. Direct costs refer to those resources that are unequivocally attributable to the object of the analysis (the drug administered to the patient or the time spent by the doctor to implant the embryos are unequivocally attributable to the IVF cycle). Indirect costs refer to the share of costs that pertain to the general functioning of the organization or the unit and that cannot be directly attributed to the cost objective (e.g. the cost of the administrative personnel or for the maintenance of the technological equipments if the cost object is the IVF cycle). To perform a full costing analysis we identified all the costs of the organizational unit delivering IVF services (the AR centre) and then we apportioned its costs to IVF treatment. In this sense, we used a top-down approach as we started from aggregated costs. However, we use detailed information collected in the two organizations two estimate direct costs attributable to IVF cycles and to identify drivers for the allocation of indirect costs.

The AR centre run by the NHS cost about € 5.2 million in 1998 (table 7.12). A part of these costs refer to services that are attributed to the AR centre but technically produced by other organizational units of the hospital. Costs for the operating room, ward care and diagnostic services are attributed to the AR centre but concern other hospital's departments. For these cost items it was possible to make a direct attribution to thethe AR centre.

The AR centre employed 8 full time equivalent workers: 2 biologists, 2 physicians, 2 secretaries and 4 nurses/technicians. The physicians and the biologists spent only a part of their working time in the AR centre. According to our estimates, total direct costs (directly attributable) amounted to € 3.7 million and indirect costs to € 1.7 million. The private centre presented lower costs for virtually all items and provided fewer IVF cycles. The overall distribution of costs is similar, although it appears less labour intensive in the private hospital. It is likely that this is due to the way personnel time is allocated to the organizational units in the two hospitals.

Table 7.12. Costs of Assisted Reproduction centre (year 1998)

Cost categories	Public Assisted Reproduction Centre (€)	Private Assisted Reproduction Centre (€)
Personnel	637,426	206,626
Drugs	118,925	68,463
Disposables and other material	199,760	89,008
Equipments (amortization)	219,317	73,630
Diagnostic services	416,588	180,657
Operating room	918,007	337,765
Hospital care	1,212,647	815,490
Total direct costs	*3,722,669*	*1,771,638*
Utilities	21,296	22,558
Other	22,908	30,500
Overheads	1,370,684	505,505
Building (amortization)	69,958	51,006
Total indirect (allocated) costs	*1,484,845*	*609,568*
Grand Total	**5,207,514**	**2,381,207**

Original data were expressed in Italian Lira; here they are reported in Euro at the parity rate of 1936.27 Liras for 1 €.

In addition to IVF services the AR centres provided other services to infertile couples and dedicated a substantial part of their resources to other units of the maternity department (table 7.13). Our informants estimated that the percentage of the time spent by the personnel of the AR centre for IVF services accounted to about 60% and 50% of the total in the NHS and in the private hospital, respectively. Similar percentages were found for drugs and disposables. However, in both centres it was estimated that most of the equipment costs had to be attributed to IVF. As operating room and hospital care costs were almost totally attributed to IVF, about 80% of total direct costs were attributed to IVF services. This percentage was used to allocate indirect costs.

Table 7.13. Parameters used to allocate costs of the Assisted Reproduction Centres

Cost categories	Estimated % of costs attributed to IVF procedures	
	Public Assisted Reproduction Centre	Private Assisted Reproduction Centre
Personnel	60	.50
Drugs	60	.50
Disposables and other material	50	.50
Equipments (amortization)	80	.70
Diagnostic services	directly allocated	directly allocated
Operating room	directly allocated	directly allocated
Hospital care	directly allocated	directly allocated
Utilities	78	83
Other	78	83
Overheads	78	83
Building (amortization)	78	83

These allocation procedures allowed us to estimate that the total costs of providing 2,257 IVF cycles in the NHS AR centre amounted to € 4.2 million (table 7.14). To provide 1,116 IVF cycles in the private centre the estimated cost amounted to € 1.9 million. These data result in a cost per IVF intervention of € 1,849 and € 1,742 in the NHS and in the private AR centre, respectively. To obtain the total cost per IVF cycle, the cost of treatment with gonadotrophin-releasing hormone (GnRH) and with follicle-stimulating hormone (FSH) need to be added to this value. Treatment costs vary across the two centres because in the private one the dosage of FSH reported in the protocol is significantly lower. Overall, including GnRH and FSH treatments, the total cost per IVF cycle amounted to € 2,732 and € 2,487, in the NHS and in the private centre, respectively. In the base case for the cost-effectiveness and cost-benefit analysis the average of these two values was used as the cost of one IVF cycle.

Table 7.14. Costs of providing In Vitro Fertilization Services and cost per IVF cycle

Cost categories	Public Assisted Reproduction Centre (€)	Private Assisted Reproduction Centre (€)
Personnel	382,456	103,313
Drugs	71,355	34,232
Disposables and other material	99,880	44,504
Equipments (amortization)	175,453	51,541
Diagnostic services	393,270	174,525
Operating room	752,327	304,711
Hospital care	1,107,806	733,373
Total direct costs	*2,982,547*	*1,446,199*
Utilities	17,062	18,414
Other	18,354	24,897
Overheads	1,098,171	412,647
Building (amortization)	56,049	41,636
Total indirect (allocated) costs	*1,189,635*	*497,594*
Grand Total	**4,172,182**	**1,943,793**
Number of IVF cycles provided in 1998	2,257	1,116
Cost of IVF per cycle	1,849	1,742
Cost of treatments with GnRH and FSH per cycle	883	745
Total cost of care per IVF cycle	**2,732**	**2,487**

On the basis of the analysis conducted by Newman et al. (1994) (see also section 5.12) we estimated that the average cost of adverse event per IVF cycle is € 522. Consequently, the total cost per IVF cycle is € 3,131 (€ 1,765 for the provision of the IVF cycle, €814 for drug treatment and €522 for the adverse events).

7.11. Cost-effectiveness of In-Vitro-Fertilisation

In the base scenario used to calculate the cost-effectiveness ratio of IVF the delivery rate per initiated cycle is estimated at 12.90% (table 7.15). Assuming that at each successive cycle the marginal probability of success declines of 1%, cumulative delivery rate is 23.3% after two cycles and 31.6% after three cycles. In other words, according to our base case almost one couple out of three will leave IVF with one (or more) baby after a maximum of three cycles.

Table 7.15. Effectiveness of IVF (marginal and cumulative delivery rates)

	Marginal probability of success			Cumulative probability of success	
	1st cycle	2nd cycle	3rd cycle	2nd cycle	3rd cycle
Base case	12.9%	11.9%	10.9%	23.3%	31.6%
Women < 40 without male infertility factor	18.8%	17.8%	16.8%	33.3%	44.5%
Women > 40 without male infertility factor	6.7%	5.7%	4.7%	12.0%	16.2%
Women < 40 with male infertility factor	14.7%	13.7%	12.7%	26.4%	35.7%
Women > 40 with male infertility factor	5.5%	4.5%	3.5%	9.8%	12.9%

The cumulative probability of leaving the service with a child greatly varies according to the age of the woman and the presence of male infertility factor. In the worst scenario (woman older than 40 with male infertility factor) the probability of success after three cycles is only 12.9%. On the contrary, younger women have much higher chances (35% or 45% according to whether the male infertility factor is present).

From an economic perspective the key question is to measure value for money that is the cost of gaining the expected outcome. As pointed out earlier, in the context of IVF an appropriate outcome measure is the delivery rate (called by some authors the maternity rate or live birth rate). Therefore, cost effectiveness analysis takes the form of the cost per delivery that is the expected cost of obtaining a “statistical” baby from IVF.

Under the assumptions of the base case, the first cycle of IVF costs € 3,131 (if the costs of side and adverse events are included) and has an estimated probability of success of 12.9%; consequently, the first cycle presents an incremental cost-effectiveness ratio of € 24,274 (table 7.16). The cost per delivery increases at successive attempts; at the second cycle is € 26,314 and at the third is € 28,728. This is because with each failed cycle the probability that the next cycle will be successful is revised downward. The raw data of the base scenario can be also used to estimate the incremental cost effectiveness ratio of an IVF programme consisting of a maximum of three cycles. A couple entering such a programme would expect to incur costs for € 8,618 and to have a probability of 31.6% to leave the service with a child. This programme would have a cost per delivery of € 27,246. These estimates were used to present cost and effectiveness information in the contingent valuation survey.

Table 7.16. Cost-effectiveness Analysis for the base scenario (World Register Data) (expected cost per delivery)

	At 1st cycle	At 2nd cycle	At 3rd cycle	After the 2nd cycle	After the 3rd cycle
Incremental costs (€)	3,131	3,131	3,131	5,859	8,618
Incremental effectiveness	12.9%	11.9%	10.9%	23.3%	31.6%
Incremental cost-effectiveness ratio (€)	24,274	26,314	28,728	25,183	27,246

Sensitivity analysis assesses the robustness of results and provides interval estimates of the measures (table 7.17). Clearly, the incremental cost effectiveness ratio is sensitive to the assumptions about the probability of success of IVF. In turn, this is very sensitive to the age of the woman. Cost per delivery ranges from € 16,000 to € 24,000 for women under 40 to more than € 70,000 for women above 40. The cost-effectiveness ratio is also sensitive to the cost of IVF. Using the range of median prices of the IVF procedures and of treatments observed in 8 Italian regions in 1998 (Mantovani et al, 1999), the cost per delivery ranges from about € 16,000 to € 67,000.

Table 7.17. Cost-effectiveness analysis: sensitivity analysis

	Incremental cost-effectiveness ratio (€)				
	At 1st cycle	At 2nd cycle	At 3rd cycle	After 2nd cycle	After 3rd cycle
Base scenario	24,274	26,314	28,728	25,183	27,246
Women < 40 without male infertility factor	16,656	17,592	18,639	17,063	18,549
Women > 40 without male infertility factor	46,737	54,937	66,626	50,366	55,753
Women < 40 with male infertility factor	21,302	22,857	24,657	21,991	23,800
Women > 40 with male infertility factor	56,935	69,587	89,469	62,451	70,335
Upper limit of cost per IVF cycle (€ 7,747)	60,053	65,100	71,072	62,302	67,405
Lower limit of cost per IVF cycle (€ 1,859)	14,413	15,624	17,057	14,952	16,177

The cost and cost-effectiveness estimates obtained in this study are higher than to those of the Italian study. Mantovani et al. (1999) estimated that the cost per delivery ranges from € 11,100 to € 19,500. These ratios derive from different assumptions about effectiveness. Cost of initial consultations, laboratory tests, egg retrieval game culturing and embryo transfers and drug treatments are very similar in the two studies (around € 3,000 per cycle). Instead, the probability of success assumed in the two studies greatly differ; while in the present study the cumulative effectiveness after three cycles is 31.6%, in the study by Mantovani et al. (1999), it ranges from 50% to 62.2% depending on the drug used for the hormone stimulation.

Estimates of effectiveness used in this study are similar to those presented by Neumann et al. (1994) and Ryan and Donaldson (1996) (about 12%). However, in these studies the cost of a successful delivery appears higher, especially in the American one. This is because the cost per initiated cycle in the US study was estimated at $ 8,000 (€ 9,528), a value which is more than three times the Italian estimate presented above. The British estimate of the cost per delivery appears close to the value obtained in this study. Based on charges to health boards who purchased IVF for the financial year 1991/1992 and including estimates for costs due to multiple births and to treating side effects for women, Ryan and Donaldson obtained that the cost per delivery (maternity) of IVF was UK £ 22,491 (€ 32,758). This value is about 20% higher than the cost per delivery obtained for Italy. Given important differences between the two countries (organization of labour, prevailing prices) and the band of variation of exchange rates this difference appears modest and seem to provide a test of convergent validity of the costing methodologies used in the two studies.

7.12. Cost-benefit analysis

First of all, cost data can be related to the WTP for the personal use of IVF. As calculated in the previous section, the expected cost of three cycles of IVF is estimated to be € 8,618 (including collateral cost of IVF as the costs of more complicated pregnancies). The cost of IVF is an issue for 53% of respondents; the remaining 47% would not try IVF even if it were free. This means that for these respondents the value of IVF is 0 (if not even negative). For these individuals the user value component of an IVF programme does not generate any welfare improvement. Out of the 3,040 individuals willing to use IVF, 2,161 were willing to pay an amount that was insufficient to cover the cost of IVF (table 7.2). In a real market, and provided that individuals acted as they stated in the survey, these respondents would be willing to pay an insufficient amount to purchase the intervention. In a real (not subsidized) market they would not use IVF. Only 879 individuals (15.3% of the total sample and almost 30% of those who would use IVF) would be willing to pay at least the cost of the procedure. Therefore, out of 100 individuals with infertility problems about 15 would use IVF if its cost had to be completely covered by consumers. Government funding would make possible the access to the procedure to an additional 38%.

Data on WTP for private use can be used to make a cost-benefit analysis that neglects benefits due to altruism. Such an analysis measures only whether total benefits of the users of the programme outweigh costs. Mean WTP for personal use of IVF is € 4,767 and € 5,474, according to the way to treat the responders who stated to be willing to pay the maximum value presented. The former is based on the assumption that maximum WTP is the value of the highest bid, while the latter is based on the extrapolation of maximum WTP. As respondent were provided with the information that IVF would have been effective with a probability of 0.3, the correct cost value to include in the analysis is that of three cycles (that has a cumulative effectiveness of 0.316). Data presented in the previous sections allow estimating the mean cost of providing three cycle of IVF to be € 8,618.

According to this special type of cost-benefit analysis, IVF would present negative net benefits: mean WTP does not cover mean cost; this is due to the fact that a large fraction of the sample would never use IVF and thus does not assign any value to the procedure. Results change if we include only respondents who are willing to pay for IVF. Mean WTP for personal use of IVF of this fraction of the sample varies from € 9,000 to € 10,381; these values are greater than the mean cost (€ 8,618). Therefore, if we include in the analysis only respondents who are willing to use IVF, benefits exceed costs.

In our opinion, these partial cost-benefit analyses should be undertaken cautiously. A full understanding of the value that people attach to IVF to overcome infertility must include altruistic considerations. Indeed, our results show that they are very important. Nevertheless, if we limit benefits to those stemming from personal use, the analysis should be limited to users, thus excluding respondents who state that they would not use IVF even if it were free. To understand if the IVF programme produces positive benefits we should compare the cost of providing the service to the benefits of its use. Consequently, mean cost of IVF should be compared to the mean WTP of users. Those who have zero WTP for IVF would not use the service and thus would not generate any cost.

From a societal perspective total benefits attributable to the programme can be measured as the aggregate WTP for the programme. If the possibility that respondents have negative WTP for the programme is ruled out, mean WTP for the IVF programme calculated from TIOLI data varies from a minimum of € 10.40 to a maximum of € 78.80, depending on the assumptions about the minimum and maximum admissible WTP that individuals can have. If it assumed that 0 is the minimum WTP and It Liras 50,000 (€ 25.80) is the maximum WTP, mean WTP varies from € 10.40 to € 19. According to the estimate derived from the MPC questions, mean WTP is € 27.30.

These estimates can be used to calculate aggregate WTP of the Italian population. Census data of year 2001 report that people over 18 years amounted to 47,162,576 units (ISTAT, 2005). If the estimates derived from the contingent valuation survey can be extrapolated to the Italian population, aggregate WTP for the IVF programme varies from € 483 million to € 1,431 Million (table 7.18).

The cost of a programme providing IVF to Italian infertile couples can be derived from the estimates of the costs of the programme (see previous section). In the questionnaire it was suggested that the programme would allow about 10,000 infertile couples to have a baby. Cost per successful delivery with IVF varies depending on the cycle (subsequent cycles have higher cost per successful delivery), the age of the woman and the cause of infertility. According to cost and effectiveness data presented in the previous chapter, the average cost per successful delivery with IVF is about It. Liras 46 million (€ 23,752). Consequently, the cost of an IVF programme that would procreate 10,000 babies from infertile couples would amount to It. Liras 460 billion (€ 238 million).

Table 7.18. Cost-benefit analysis of a programme providing IVF to Italian infertile couples (10,000 babies born from IVF)

Assumptions on WTP estimates	Mean WTP per survey respondents (€)	Aggregate WTP (benefits in million €)	Total costs of the programme (costs in million €)	Net benefits (million €)
TIOLI data (model 1)	19.04	898.08	237.52	660.56
TIOLI data (model 2)	16.96	799.88	237.52	562.36
TIOLI data (model 3)	14.16	667.82	237.52	430.30
TIOLI data (model 4)	10.39	490.02	237.52	252.50
Modified payment card data (un-weighted)	27.28	1,286.32	237.52	1,048,80
TIOLI data, Logit no spike (model 1)	23.60	1,113.04	237.52	875.52
TIOLI data, Logit with spike (model 1)	26.00	1,226,27	237.52	988.71
TIOLI data, Logit no spike (model 2)	10.25	483.42	237.52	245.90
TIOLI data, Logit with spike (model 2)	30.34	1,430.91	237.52	1193.40
Modifie payment card data (weighted)	26.59	1,253.79	237.52	1,016,27

For all the scenarios total benefits greatly exceed total costs. Even if WTP is derived from a very conservative hypothesis which assumes that only "yes, definitely" responses reveals WTP and all other possible responses (including "yes, probably" and "do not know") reveal that respondents are not willing to pay, aggregate benefits are more than the double of total costs. In the most conservative scenario, that assumes that only respondents with "yes, definitely" answers are willing to pay for the programme costs are about 50% of benefits. Only if we assume that a substantial fraction of respondents have negative WTP the programme results as having negative benefits.

According to cost-benefit analysis principles, a programme makes a positive contribution to the welfare of the population if benefits expressed in monetary terms exceed costs. It is thus possible to calculate the minimum value of mean WTP that would suffice to cover the costs of the programme. This corresponds to total costs divided by the number of individuals being part of the population (in this case the adult Italian population). About a mean WTP of € 5.04 per member of the population would suffice to cover the cost of an IVF programme generating 10,000 babies from infertile couples.

TIOLI data may be used as they resulted from a real referendum. As mentioned earlier, if a strict majority rule applied (absolute majority of voters), a tax of It. Liras 10,000 (€ 5.20) to fund IVF would have been approved. Had a less stringent rule adopted "voters" would have approved a higher tax. If a relative majority sufficed to approve the proposal (the number of "yes" greater than the number of "no") a tax of It. Liras 20,000 (€ 10.30) would have been approved.

Welfare estimates (those based on mean WTP) and decisions based on majority rules applied to TIOLI data are in favour of publicly funding IVF to infertile couples. Overall, our cost-benefit analysis of IVF in Italy for couples with infertile problems shows that it would imply a welfare gain.

Chapter 8

Conclusions

8.1. Introduction

The research presented in this thesis makes a number of main contributions, some methodological and some substantive. The methodological contributions concern the feasibility and the validity of cost-benefit analysis and contingent valuation in the area of reproductive medicine. Cost-benefit analysis plays a pivotal role in normative economics as it is expected to offer guidance to decision makers about the allocation of economic resources. Intuitively, doing a cost-benefit analysis means clearly identifying the object of the evaluation (the programme) and to compare costs (use of scarce resources) to benefits (contribution to the wellbeing of people) attributable to the programme; when benefits exceed costs the programme is valuable, and thus it is worth being funded. Not surprisingly, there are a large variety of techniques which are consistent with this basic idea of cost-benefit analysis. National governments, international institutions (e.g. UN agencies) and organizations (e.g. World Bank) and non governmental organizations have their own way of doing cost-benefit analysis. Although these methodologies may significantly vary in several respects, what they have in common is their attempt to use a logical framework to provide actual guidance to decision-making.

Our research started from a similar stance. We wanted to use cost-benefit analysis to provide an answer to a precise policy question. The question was: Should the Italian government fund IVF treatment for infertile couples? This policy issue has driven the choice of the conceptual framework, the research design and the investigation methods. Within this policy perspective, in the first part of this concluding chapter we review the major methodological issues addressed in the research and provide some critical insights.

The other major contribution is strictly connected to the methodology and refers to the policy implications of the results of the study. Overall, results support the inclusion of IVF (and to a certain extent of other Assisted Reproductive

Technologies) in the basic package offered by the Italian National Health Service. This is an important outcome of the research and it is worth discussing in the light of the recent Italian policy of being more transparent and explicit in the rationing process, and in the light of the regulation of reproductive medicine. In addition, results of the study have implications beyond the decision of whether to include IVF in the national guarantees as they help understanding how Italian citizen view reproductive medicine and the place it should have in the healthcare system in general. These themes are discussed in the second part of this chapter.

In summary the thesis shows that:

- An IVF programmme providing free IVF to infertile couples implies a welfare improvement for the Italian society because benefits (€ 400-1,300 millions) exceed costs (€ 240 million);

- Altruism is an important driver of this result as respondents are willing to pay substantial amounts of money even if they do not need IVF and, more importantly, even if they would never personally use IVF in a hypothetical situation of infertility;

- Societal benefits are mainly attributable to a minory of respondents with very high WTP; while cost-benefit analysis fully appreciates the intensity of preferences, democratic processes (like referendums) do not; according to our survey only 55% of Italians would have approved a tax of € 5.20 hypothecated to a national programme costing € 5.02 per person.

8.2. Testing actionable cost-benefit analysis

Governments and other organizations acting for the public interest need to be guided by economic methods. Clearly, in order to be useful these methods have to be feasible and valid. They have to be feasible in that it has to be possible to conduct studies in the context of actual decision-making. Typically, in real contexts resources and time are limited. Validity refers "to the extent of matching, congruence, or "goodness of fit" between an operational definition and the concept it is purported to measure" (Singleton and Straits, 1993: 114). In other words, the method is valid if it really measures what it intends to measure. The main challenge of the thesis was to measure the net benefit of providing IVF treatment to infertile couples. Consequently, to discuss the validity of the method requires checking whether it really measured the net benefit of such a programme. In turn, to establish the validity of the cost-benefit method used in this study a conceptual definition of net benefit is required. Without an appropriate theoretical framework, it would be difficult (if not impossible) to assess the validity of the method. The theoretical framework adopted in the study is that of welfare economics.

Although it would be possible to have a physical measure of the benefits of IVF (that is births or pregnancies), we preferred using the notion of value (Williams, 1983) and to refer to the willingness-to-pay approach. This

> "*..approach is derived directly from welfare economics theory and hence is generally recognized to be conceptually appropriate for establishing individuals' values from a welfarist perspective*" (Birch and Donaldson, 2003: 1121).

A number of authors argue that this should be the preferred approach to be used for evaluation health care programmes (Pauly, 1995; Johannesson, 1996a; Birch and Donaldson, 2003). In addition, in the area of reproductive medicine other economic methods may be inappropriate or of limited scope. It would be possible to use cost-effectiveness analysis to calculate the cost per life gained (generated) of IVF. Indeed, in Chapter 7 we present this calculation. However, cost-effectiveness analysis of IVF cannot be used to make allocative choices because the benefit of a new life cannot be compared to the cost of a life saved. The outcome of IVF and other ARTs is such that cost-effectiveness analysis is of little help to decide government funding.

Cost-utility analysis is increasingly gaining acceptance in the health care sector. While it tries to capture both quality and quantity of life gains of health services, it also tries to estimates "benefits" consistently with the axioms of utility theory under uncertainty. In theory, cost-utility analysis could be used to assess the benefits of IVF. In practice, however, it would be very difficult to elicit the utility of an IVF baby through standard-gamble and time-trade-off exercises and, as far as we know, it has never been attempted. Both cost-effectiveness and cost-utility analysis do not appear the appropriate methods to assess the economic value of IVF. The appropriate analysis to measure benefits of IVF is cost-benefit analysis that is the analysis where the value of the service is measured in monetary terms and compared to its cost.

8.2.1. How to measure benefits of an IVF programme

The theoretical base for the measurement of benefits is the concept of consumer surplus (Hicks, 1941; Johannesson and Johnson 1990). In principle, the measurement of consumer surplus can make use of the willingness-to-pay or the willingness-to-accept approach. In the former case the maximum that an individual is willing to pay in order to obtain the good is measured, in the latter the minimum he/she is willing to accept as a compensation for losing the good is measured. A major problem of theoretically grounded cost-benefit analysis is how to estimate WTP or WTA. One possibility is to make inferences from actual behaviour of individuals (revealed preferences). However, for many goods, especially those for which we are interested in performing cost-benefit analysis, actual transactions do not exist or do not take place in competitive markets so that market signals are lacking, distorted or difficult to capture and interpret.

Given its flexibility, the CV method can be used to measure WTP or WTA. In practice, however, WTA is rarely used when the respondent does not have the good being valued. In such a case the question is cognitively difficult because the respondents should first imagine having the good and then they should formulate the compensation of having it taken away. The willingness to pay questions are easier to understand, especially if the respondent is not familiar with the good. Therefore, we performed the cost-benefit analysis making WTP questions aimed at measuring the compensating variation (the amount of money willing to pay for the benefit) for an IVF programme.

The CV method is a powerful approach to be used for priority setting and to offer guidance to the allocation of resources in publicly funded health care systems. Depending on the decision-making context, various approaches can be followed. A major choice concerns whether one programme at a time is evaluated or several programme simultaneously are evaluated (Olsen et al., 2005). Both approaches present strengths and weaknesses. Although a thorough theoretical discussion of each of them is beyond the scope of this thesis, it is important to outline some major methodological issues.

On one hand, the focus on a single programme allows the respondent to be more informed about the programme. In addition, this approach mimics purchase decisions and voting in referendum, in that in real life decisions regard one thing at a time (even if people make trade off in their minds). On the other hand, this approach has been criticized by Olsen and Smith (2001) to be of no aid to decision-makers if the healthcare system operates under a global budget.

It is argued that even if WTP for a new programme exceeds its marginal social costs, it is not guaranteed that it provides a welfare improvement. Welfare gains require that marginal benefit equals marginal cost for all existing programmes and this is unknown unless all health services already funded are investigated. In addition, if budget is fixed it is necessary to rank programmes to the extent of their net benefit so that total net benefits are maximised under the funding constraint (that may insufficient to fund all services with benefits greater than costs). Hence, Olsen and Smith (2001) conclude that: "a partial valuation is not an aid to decision-makers, if money is to be taken from a given health care budget". This critique is a serious concern for cost-benefit analysis and it is a thorny issue in welfare economics as it calls for "second best" solutions.

In this study we did not frame the WTP scenario and questions on the basis of the assumptions that the budget is given. Indeed, it was clearly stated in the questionnaire that additional taxes were raised to fund the IVF programme. If the welfare function is separable in respects of the programmes funded by the given budget, cost-benefit analysis of one programme at a time may provide useful results. If the WTP for a new programme exceeds its marginal social costs and its provision is funded outside the given budget, the addition of the new programme improves social welfare. Notice that the opposite is not true. If WTP does not exceed social costs it cannot be assumed that the programme has to be rejected. Furthermore, there is the possibility that some programmes are inefficient and would thus make it possible to free resources that would have been used for the rejected programme.

If it is possible to add resources for new programmes, evaluating one programme at a time is theoretically sound. In addition, strategy of evaluating more programmes together may be hindered by practical concerns. It is difficult to survey respondents on different programmes at the same time. An adequate description of the programme being evaluated is essential to make the survey valid, especially if many respondents are not familiar with the good that is being evaluated as it is the case with IVF. Had we described one or two programmes in addition to IVF, in the survey we would have probably reduced the attention and comprehension of respondents of the IVF programme.

As discussed in Chapter 3, WTP estimates of health benefits are often made through contingent valuation (CV) surveys where direct WTP questions asked. More recently a new approach, called Discrete Choice Experiments (DCE), has been used for estimating benefits of health care programmes (McIntosh et al., 1999; Ryan, 2004). Here respondents are asked to choose between alternatives, described in terms of a graduated set of attributes (including price). Then, econometric methods (variations of logit and probit models) are used to estimate WTP for marginal changes in attributes and overall welfare measures (Hanley et al., 2003).

The use of the DCE approach in health care is still experimental and debated (Bryan and Dolan, 2004; Cairns et al., 2002; Cairns and van der Pol, 2004; Lancsar and Donaldson, 2005). One major issue concerns the cognitive burden associated with a full DCE survey as it must include several questions with different levels of the attributes. We believe it would have been too cognitively demanding to conduct a WTP survey with several questions having different levels of the attributes of the IVF programme. The "traditional" CV approach was deemed more viable. It is worth noting that one of the very few studies comparing the two methods is on ARTs. Ryan (2004) conducted two studies with the same users of IVF in an Assisted Reproductive Unit at Aberdeen and found that the welfare estimates derived from the two methods were not significantly different.

Our survey provides some evidence that contingent valuation is feasible and valid. This is the first large CV study in Italy and the first attempt to survey WTP for IVF of a sample of the general population. Other studies (Ryan 1996, 1997, 1998 and 2004; Neumann and Johannesson, 1994) have either surveyed people attending ART centres or convenient samples of students and health professionals. Our survey involved individuals of all ages and of a large variety of socio-economic conditions. The low number of inconsistent answers, the extremely low number of answers that can be considered as "protest" answers, the positive correlation between socio-economic measures and WTP and the consistency of results across elicitation methods are all elements in favour of the validity of the survey. This study shows that asking the general population WTP questions is feasible and provides internally valid answers.

A widely debated issue in CV research concerns the elicitation method. The guidelines of the National Oceanic and Atmospheric Administration (NOAA) of the US government (Arrow et al., 1993) have had a strong influence in this respect, including our research. They strongly argued in favour of a referendum format where respondents are asked to answer yes or no and are allowed to refuse to answer. We followed this recommendation as we randomly split our random sample in 4 groups and to each we submitted a different value. In addition, we included the "do not know" option. We also articulated both "yes" and "no" answers to get additional information on the strength of preferences. Respondents could choose between probably and definitely ("yes" or "no"). Given that we had a tight constraint on the number of questions, we did not follow the standard of framing a "yes"/"no" question first, and then asked respondents how much they were sure about their answers. Rather, we opted for only one question with 5 options ("yes, definitely", "yes, probably," "no, probably", "no, definitely" and "do not know"). We are aware that the two approaches are not equivalent as they are associated with different cognitive processes. However, the relative advantages and disadvantages of the various approaches to sequence questions are presently unclear, at least to our knowledge.

As we were aware that to use only four values would have produced estimates with large confidence intervals for the entire sample we included a subsequent series of yes/no questions to a larger number of values. While this method is similar to the MPC as far as the calculations to estimate WTP are concerned, it is not a proper MPC because it presented 8 different specific values and asked respondents to state yes or not to each one. The list of values included the four values used for the referendum format and other four values so to offer respondents the possibility to choose from It Liras 2,000 (€ 1) to It Liras 200,000 (€ 103). Hence it was possible to compare the answers of the two questions' formats and to undertake a convergence validity test. Obviously, a major limitation of this check is that it cannot be assumed *a priori* that the answers of the questions are independent.

8.2.2. Eliciting WTP for an IVF programme: Whose value?

Although IVF appears a popular topic for the CV literature, probably because of the limitations of cost-effectiveness studies in the area of reproductive medicine, we are aware of only two studies that draw clear policy conclusions from a cost-benefit analysis of IVF (Granberg et al., 1995; Ryan, 1997). These studies were discussed earlier (Chapter 2 and 3); it is worth mentioning here that they are based on surveying IVF users and thus that policy implications do not derive from the perspective of those funding such a programme in a public system.

Granberg et al. (1995) concluded that, according to their study conducted in one centre in Goteborg, the benefit to the infertile couples (those treated in the Centre) was higher than the cost to the Swedish NHS. They did not state that this result implied that public funding was warranted, although it is implicitly suggested. On the basis of cost data obtained in Scotland (Ryan and Donaldson, 1996) and WTP data from 466 IVF users of the Aberdeen's Assisted Reproduction Unit (Ryan, 1997; Ryan, 1999) a more explicit conclusion was drawn: "The results suggest that the benefits of providing the service outweigh the costs, and that public provision of the service should be encouraged" (Ryan, 1997: 842).

Our survey asked respondents their WTP for personal use of IVF. This way we measured the hypothetical user value of IVF and could obtain a measure partly comparable to that obtained by the above-mentioned studies. Obviously, a major difference is that our respondents were asked to imagine a hypothetical situation, while in Granberg et al (1995) and Ryan (1997) studied patients who actually experienced infertility and went through IVF. In both surveys it was performed an ex-ante evaluation of the programme with the aim of measuring the user value of IVF. Hence, our survey allows making a cost-benefit analysis similar to that proposed by Ryan, but based on the preferences of the general population rather than of the actual users of IVF. Although hypothetical situations may be associated with special validity and reliability problems, this approach allows the measure of welfare gains from the point of view of those who are requested to pay for the publicly funded programme.

In our sample, it would have been useful to compare WTP between those who experienced infertility and IVF those who did not. However, no question was asked about the personal infertility problems and the use of ARTs. Such questions were not asked to avoid making the survey too sensitive to emotional issues. Ex-post, we think that we should have included a specific question concerning the personal experience of the respondent with infertility and IVF. However, it is worth noting that the two questions asking for the acquaintance of respondents with people with infertility problems or who experienced IVF may serve, to a certain extent, as proxies.

Whilst there is no doubt that user value is an argument of the utility function of individuals, the issue of whether other arguments may be included is still debated. Birch and Donaldson (2003, 1122), following Culyer (1991), argue that assuming that " i) social welfare is a function of individual utilities and ii) individual utilities are a function of the commodities (i.e. goods and services) consumed by those individuals" are restrictive assumptions that, at least partly, can be relaxed without violating the general axioms of welfare economics. The issue of values other than those derived from the personal use of the good has been discussed in health economics in the context of whose values CV studies should be elicited (Hanley et al., 2003). Two other sources of value are deemed relevant: "caring externality" (utility derived from knowing that someone else is using the good) (Culyer, 1976; Ryan, 1996; McIntosh, 1999; Hanley et al. 2003) and "option value" (utility derived from knowing that a good is available for future use given demand uncertainty) (Weisbrod, 1964; Ryan, 1996; McIntosh, 1999; Hanley et al. 2003).

The "caring externality" reflects an altruistic feeling while "option value" reflects a personal interest in having a sure access to a good whose demand is uncertain, and it is associated to the value of an insurance product. Both these non-user values can be captured in a survey of the general population if the programme being evaluated is the provision of IVF services to infertile couples with public funding.

In this respect our study provides new evidence. We used a national sample to elicit total WTP for a programme providing IVF to infertile couples with public funds and, at the same time, WTP for a hypothetical personal use of the service. Our estimates of WTP for personal use of IVF is not a component of total WTP because survey panellists were presented with an *ex-post* scenario in which they had to imagine being infertile and willing to have a baby. If we had only formulated the demand as a probabilistic *ex-ante* scenario, WTP for using IVF would have had to be smaller than total WTP. Nevertheless, our study provides new evidence on the existence of a pure altruistic component of WTP. About two-thirds of respondents who would have not used IVF were willing to pay something for a publicly funded programme. WTP of these individuals is purely altruistic because the programme would not be used by the respondents even if it were free. It is worth noting that these respondents stated that they were not willing to pay in a hypothetical situation. For them, altruism goes beyond the concept of fair innings. They would be paying to make something available to others that holds no value for them, even in a hypothetical situation. This finding is new in the area of IVF and ARTs and shows that procreation is conceived by people as an area for solidarity of funding arrangements.

The particular way we framed the questions allowed us to make a distinction between two kinds of altruism that may deserve attention in future studies. Normal altruism refers to the willing to pay for something that is not used but would have been used by the subject if he/she were in a different condition. Pure altruism refers to the situation where a person is willing to pay for something he/she would never use under any condition. This form of altruism reflects a combination of solidarity and freedom, as it expresses the willingness of individuals to help others to make different choices. It seems to us that donating in order to let others have the same goods we have is profoundly different from donating to let others have a good we do not want to have. The latter, at least in some circumstances, reveals a larger scope for solidarity.

8.2.3. Eliciting WTP for an IVF programme: using Internet

The idea of surveying the general public on IVF was tested in a study performed in Boston at the Harvard School of Public Health (Neumann and Johannesson, 1994). The authors used a convenience (clearly unrepresentative) sample and consequently did not draw policy conclusions from their analysis. They also clearly stated with reference to IVF that "to use results for policymaking, it will be important to use a representative sample of the population in future studies".

From the beginning our study wanted to provide a picture of the WTP for an IVF programme of an entire population, so to provide an analysis that could really offer guidance to policy making. We reviewed the major methods of administration of CV questionnaires and found that they were either inappropriate (it would be difficult to present IVF and to make WTP questions in a telephone interview) or too costly (face-to-face interviews are appropriate but were unaffordable given the budget). These constraints led us to test a new method to administer CV surveys: internet. When we designed the survey no scientific literature was available to understand its pros and cons and very few scientific studies had been conducted. Nevertheless, this method was appealing to us because it was seen innovative, and thus represented an opportunity for testing a new method, and was affordable.

We described how the survey worked and the characteristics of the panel that we used in Chapter 5. Here we simply summarise the potential benefits of the method: it reaches high response rates, it introduces flexibility in the questionnaire design and eases the control of respondents' and sample characteristics. In our opinion these advantages justified the test presented in this thesis.

Our survey performed well and gave results consistent with surveys administered by telephone, face-to-face or by post. The number of respondents whose answers were inconsistent is limited (3-5%) and the number of answers that could be clearly detected as "protest" was about 20 (about 0.5%). It appears that the survey panel took the questionnaire seriously. Results of regression analyses appear consistent with *a priori* expectations. More importantly, the income of the respondent (or the self-attributed socio-economic class) is positively associated to WTP for IVF and, in the referendum format the higher the value offered the smaller the fraction of respondents willing to pay. These basic validity checks show that this method to administer the survey did not present any special problem and performed similarly to the others.

Further evidence in favour of this new method of collecting CV data come from two problems often encountered in the CV literature, that have also been detected in this survey . First, the WTP estimates derived from the referendum format exceeded the estimates derived from the modified payment card (MPC) format. Second, we detected a relevant anchor biases in the WTP answers. Respondents appear influenced by the value given in the referendum format when answering to the MPC questions. Both these problems are largely known in the CV literature. To detect them in a large internet based survey confirm their presence across any type of survey method.

Overall, this study presents important evidence in favour of using internet to survey people for eliciting WTP. Despite the lack of experience in designing internet surveys and little help from the literature, it appears that the survey worked well and that respondents had "normal" reactions to the stimuli of this method of administration of the questionnaire. We could not detect anything seriously abnormal in the survey. We do think that this attempt to use internet to elicit WTP for a health care programme has been positive. Given the potential benefits of the method, internet surveys should deserve more attention by the health economics scientific community.

At the moment, an internet survey cannot be fully representative of the general population. The older and less educated are seriously underrepresented in the sub-population of internet users. This is a serious issue that may undermine the utility of the method. However, age and education may introduce biases also with the other methods and, given the large amount of information that can be collected on survey panellists, post survey corrections can be used to make the sample more representative. In effect, we adjusted estimates obtained from the survey to make it more representative of the Italian population.

Finally, we also want to highlight that the use of an electronic panel established and administered by an independent institution significantly limited the room we had to design our study. We had a restricted number of questions that we could administer, we had to avoid highly sensitive questions and, more importantly, we had very limited room to produce different versions of the questionnaire to submit to sub-samples. In this respect we paid a relevant price for the use of this novel method of survey administration. The possibility to collect data from a sample of the entire Italian population costed us a sub-optimal design of the questionnaire. Having said that however, we want to underline that CV is a sophisticated and complex technique to gather relevant information and that, unavoidably, conducting a good study means to decide about trade-offs and to find a balance between different instance that contribute to the quality of the study (sampling, framing of the questions, methodological tests).

8.2.4. Eliciting WTP for an IVF programme: Which elicitation method?

In many respects we tried to follow the NOAA panel recommendations for conducting contingent valuation studies of environmental goods. We tried to follow these recommendations also in respect of the elicitation format. For environmental goods, the NOAA panel strongly criticized the use of open-ended questions and recommended that "the valuation questions should be posed as a vote on a referendum" (Arrow et al., 1993). We framed the first question concerning the IVF programme for IVF exactly according to these guidelines. However, we were allowed to split the sample into only four groups. Given the size of the sample, this made precise estimates for each value but collected limited information about how WTP varies for different values. Indeed we only collected information on whether respondents were willing to pay four values: approximately € 2.60, €5.30, €10, and €25.80. Results clearly show that our measurement instrument was not sensitive enough; the difference in overall "yes" answers was only 13% between the lowest and the highest bid. As we anticipated this result, we added a modified payment card (MPC) with 8 different values to all respondents. In effects, the MPC method was more informative as for the highest bid (€ 103.30) "yes" were 55.6% and for the lowest bid "yes" were 10.3%. A similar degree of sensitivity was found in the survey for personal use of IVF.

Overall both the elicitation methods that we used appeared feasible and easily understandable to the respondents. In the referendum format "yes, definitely" answers were strictly declining, but "yes" answers (sum of "definitely" and "probably") where not strictly declining because the fraction of "yes" respondents for Italian Liras 5,000 was lower that the fraction of "yes" for Italian Liras 10,000 (about 1% difference). In the modified payment card question, where values where presented randomly to each respondent, we found a limited number of inconsistencies and a very few answers that may be considered "protest".

The main purpose of this survey was not to test the CV method in general and thus it cannot provide strong arguments in favour of its validity and reliability. Nevertheless, it gave apparently reasonable results and did not show any strong evidence supporting the claim the method had serious validity problems. Really, the use of two elicitation methods in the same survey provided material to test the convergent validity of the methods and to detect anchor biases. The evidence on convergent validity is equivocal. The two methods produce different WTP estimates. If the possibility of negative WTP is excluded, TIOLI estimates results higher than the MPC estimates. Consistently with other studies (Ryan, 2004) and the position of the NOAA panel (Arrow et al., 1993), we found that TIOLI may overstate true WTP. However, surprisingly we found that the answers of the questions in the two formats were strikingly consistent. The elicitation methods are not convergent but respondents are consistent in providing their answers across the two methods. This suggests that the difference in WTP values may derive from the assumption used to make the estimates and, particularly, those determining the WTP of the respondents with the highest WTP. Definitely, this issue deserves more attention in future specific studies.

The evidence in favour of an anchoring effect is solid. Respondents are significantly influenced by an initial value, even if it stated in a different question. The direction of this bias is expected: the higher the initial value, the higher the estimate derived from the following MPC questions. Is this bias disturbing? Yes, to a certain extent it is. It would be much better to have a measurement system that it is insensitive to contingent elements. However, it is naïve to think that such a method can exist. A major learning that economists got from psychology is that preferences are responses to stimuli and hence cannot be investigated as they were an object independent by the context and the observer.

Social sciences refer to two main concepts of validity: criterion-related validity and content validity (Straits and Singleton, 1999). We have evidence on both concepts in our study. Criterion-related validity concerns the degree to which the concept under consideration enables one to predict the value of some other concepts that constitutes the criterion. In this study criterion-related validity is provided by the association between answers to the WTP questions. Construct validity is to be tested at theoretical level and requires a consistent pattern of relationships. Our study provides evidence in this respect in that some variables have the expected association with WTP. First of all, income and socio-economic status measures are significant regressors of WTP in virtually all models. Once the effects of other variables are controlled, people with higher income (or better socio-economic status) have higher WTP. This is a standard requirement in WTP studies that is fully respected in this study. In addition, most of the variables that were expected to explain variability of WTP have statistical significance and present the expected sign. Respondents who are employed, are more educated and live in larger municipalities show higher WTP for IVF, even after controlling for income. These results are consistent with expectations as IVF is a sophisticated technology that requires knowledge to be comprehended and that help couples who may have delayed procreation because of studies (Italians complete university studies later in their life than other Europeans) and professional expectations. To a certain extent, IVF and other ARTs are goods for the middle urban classes, rather than for the very rich (difficult to fully represent in such a survey) and the lowest social classes. When the models included all these variables, it also appeared that southern respondents are different from northern ones in that the former have higher WTP. Other things being equal, it appears that southern people's preference for IVF is stronger. This result is expected as it is consistent with southern values and traditions which favour large families and tend to blame childless couples.

In the case of IVF a relevant part of the population may attribute zero value to a public programme. Some individuals simply do not see any benefit in the programme and thus they have a zero WTP. The possibility of zero WTP is particularly important when the TIOLI format is used. Basically, in this method the answer to the yes/no question reveals minimum/maximum values of the WTP distribution. For example, if the lower value offered is € 5, no-respondents are only signalling that are not willing to pay that amount. How this information is used to estimate the WTP distribution or only a value of the central tendency depend on the assumptions about the probability distribution function, if a parametric or a semi-parametric approach is followed, or the algorithm used to calculate mean or median WTP if a non parametric approach is followed. The crucial point here is that the issue of zero WTP may require using a semi-parametric approach so to better model the characteristics of the preferences to be elicited. In the thesis we tested a particular semi-parametric model labelled "spike-model". Simply, the model is based on the combination of two probability density functions, one concerning whether the individual has zero or positive WTP, the other reflecting the probability that individuals with positive WTP accept different values. Intuitively, this method assigns zero WTP to a part of the sample and, compared to a "normal" logit model, preclude negative observations.

Results of the various models we used are consistent with expectations and confirm that the assumptions about the probability distribution function significantly matter. The case of IVF clearly suggests that it really matters whether it is assumed that individuals may have negative WTP for the programme. Provided that negative WTP is excluded, whether a spike or another model is used does not have any large effect because the left tail of the WTP distribution contributes little to the mean WTP. The relevant issue is how to model the right tail of the distribution because the mean WTP is substantially determined by the (relative) modest part of the sample that is willing to pay the highest amount.

The survey presented in the thesis was designed as a part of a full actionable cost-benefit analysis. Methodological choices were mainly based in the attempt to follow best practices. Nevertheless, it provides some important contributions that are worth being summarised. First, this was the first survey administered to a sample of the general population on infertility and IVF. It proved to be feasible and results appear valid. This shows that contingent valuations, even on a critical issue as IVF in Italy, can be conducted on samples of the general population. Second, a few methodological tests were made possible by the survey design. All of them appear consistent with previous studies. This is very important because they provide evidence in favour of the major methodological contribution of our survey. This is the first attempt that we are aware of that estimated WTP from the general population through an internet survey. We have collected evidence that the internet may be a valid mode to administer CV survey. This result is of paramount importance because an internet survey may present several advantages over the traditional methods, including costs, time, flexibility and control over respondents.

8.2.5. Costing methods

Although the fact that both costs and benefits have similar impact of results of cost-benefit analysis is rather obvious, researchers often pay very little attention on how to estimate costs. In this study we spent a fair amount of time and energy to produce good estimates of costs of providing IVF in Italy. First, we selected the procedure to calculate costs on the basis of a review of the possible methods available and of a basic understanding of cost accounting, the management discipline dealing with costs and their use for decision making. Second, we selected two Assisted Reproduction centres and performed a micro-costing study. Basically, we collected data and information about the two centres and used them to estimate the full costs of an IVF cycle. Therefore, our cost-benefit analysis is based on a specific costing study, based on pondered methodological choices.

In Chapter 4 we argue that full costing is the most appropriate procedure when cost-benefit analysis should inform the inclusion of an item in the basic package of the services to be provided with public funding. We think that this is the appropriate approach because it approximates long-run marginal costs. In the long run, if the decision is whether or not to provide the service with public funds, all categories of costs tend to vary and thus appropriate measures of costs should include a fair share of overheads.

In the empirical part of the study we were consistent with the full costing methodology, despite the difficulties of obtaining appropriate data. Therefore, this study was based on the collection of original data on both sides of the economic analysis (benefits and costs). As we have pointed out that costing is often overlooked in cost-benefit analysis, we think that a major contribution of this study is the illustration of an example of how data, generally available in health care organisations, can be used to estimate costs for cost-benefit studies.

8.3. Limitations of the study

Had we the opportunity to re-start the project we would make some different methodological choices. First, testing and piloting was insufficient. We were under time pressure because the use of the ACNielsen-ISPO panel was possible in a certain period only and could not delayed. For this reason we tested a hard copy of the questionnaire with a small convenience sample of colleagues and three experienced professional pollsters provided some useful comments. Had we carried out a better pilot we would have probably used more informative bids in the referendum questions and more appropriate values for modified payment card questions. Better planning could have allowed us to improve the survey. However, the issue of piloting internet survey through commercial panels is particularly critical as costs of pilots may be relevant, compared to the overall costs of the survey, and may interfere with the sampling procedures. Having said that however, after a few years since the survey was conducted the scenario descriptions appear clear and effective (see Appendix) and we do not have serious regrets about how we framed the questions.

Second, in addition to performing a full cost-benefit analysis we could have investigated more methodological issues. The substantive aim was overriding and distracted us from using the survey to test some methodological issues that were unexplored when the survey was designed (autumn 2000) and are still debated. These issues include sensitivity to the scope of the programme (e.g. by providing them different scenario in terms of beneficiaries and/or probability of success), sequencing of questions and a better design of the relationship between WTP for personal IVF use and WTP for the publicly funded programme.

Third, there are other two specific choices that may be disputed. We did not randomly select the bids for the referendum questions as we did not randomly select the values for the MPC questions. Random selection of bids was not recommended by the NOAA panel but has been used in a few studies, including one on IVF (Ryan, 2004). Lack of randomisation of bids was also found to be source of relevant biases. Our bids and values were purposely selected. For the referendum format, we think that if we had done a random selection we could have selected bids of limited informative values (e.g. to covering a limited range). The main problem in this respect was the limited number of bids that we used in the referendum format, rather than the way we selected the value. It is worth reminding that this is due to lack of flexibility in managing the electronic panel and to assure that each sub-sample was representative of the Italian population. Nevertheless, we could have randomly selected the values used in the MPC questions. The other contestable choice regards the lack of any follow-up questions. We were aware that this is an important limitation but, again, the inclusion of an open-ended question in the electronic questionnaire was not possible for technical reasons.

As mentioned earlier we did not include any question on personal experience of surveyed individuals with infertility and IVF services because they were deemed too sensitive. Probably, the inclusion of such questions would have caused some concern to a few respondents and, maybe, would have induced more protest answers. However, information about the direct experience of respondents with infertility and IVF would have substantially enriched the survey and would have made possible to investigate WTP for personal use of IVF services.

The fifth limitation is probably the most important. We did not address the issue of negative values. That is, we did not expect the possibility of hostility towards IVF. This issue was not addressed in the past, at least for what we know, and was not sufficiently addressed in this study. Really, this is probably the most relevant issue differentiating IVF from other health care programmes; it deserves attention from a theoretical, methodological perspective as it also relevant from policy perspectives. In many phases of this investigation we felt it would have been appropriate to analyze the possibility and the implications of measuring negative WTP for IVF, but we could not adequately frame the issue and we did not include any question on this issue. Ex-post, we regret not being more courageous in this respect. We strongly encourage research on negative WTP values because we do think that if we include caring externalities to capture altruistic feelings we should also investigate whether funding programmes that are deemed against personal values may have a negative impact on individuals' utilities and thus social welfare.

8.4. Policy implications

This is the first study that tried to provide an answer to the question "IVF should be publicly funded?" by investigating WTP of a national representative sample. Results clearly provide a positive answer to the question. Our WTP estimates per Italian resident, ranging from €10 to € 70, are much higher than the expected per capita costs of providing IVF to 30,000 couples to procreate 10,000 babies (between € 4.70 and € 5.20). Our results appear robust as they hold under several circumstances. As stated above, the only major challenge to our net benefit estimate is that our results hold but we do not admit that some people may experience disutility due to public funding of technologies for assisting reproduction.

This study was designed to capture the perspective of Italians. There are no doubts that countries may have relevant specificities in respect of ARTs, as they touch critical cultural, social, religious and ethical themes. Overall, however, we think that these results can have some value for other European countries. After all, religious arguments are very important in this area of medicine and the fact that Italy is a catholic country with an influential role of the Roman Church should bias against rather than in favour of public funding of IVF.

Interestingly, while we used a referendum format to elicit WTP for publicly funding IVF (the questionnaire was administered in year 2000), in June 2005 Italy had a popular referendum to abolish the new legislation restricting the use of ARTs. Reproductive medicine is one of the most divisive issues that Italian society has experienced since the end of WWII.

Legislation permitting and regulating abortion was passed by the Parliament and resisted a popular referendum in the early 80s. The Catholic Church, fiercely against this legislation, obtained the support of the largest Italian political party (Christian Democrats). The fact that about 80% of Italians went to the polls and a neat (although limited) majority voted for maintaining the legislation had important implications for the Italian society and politics. In the following 20 years ARTs was the new terrain of confrontation between the Catholic Church and a large variety of social and political groups. Up until 2005 Italy had been without relevant legislation about ARTs. Several attempts failed in the 90s. Once, despite a bi-partisan agreement, legislation that was killed by parliament members that claimed that for such a sensitive issue they had to follow their moral values and thus could not obey to the indications of their parties. Only in 2004 was a national law on ARTs approved by the Parliament. The law was passed because of the large majority of the Berlusconi government in both branches of the Italian Parliament and the strong support of the Catholic Church. The new legislation permits IVF and other ARTs, but only with couples' genetic material. In addition to forbidding heterologous ARTs, the legislation bans freezing embryos and limits the number of embryos that can be implanted in the woman's womb to three. This is one the most restrictive legislation that major European countries have on ARTs and is reported to cause migrations to other countries to circumvent it.

The national referendum, called by two million Italians who wished to abolish this legislation, failed because only 34% of adult Italians (automatically registered to vote) went to the polls. Basically, those who favoured the approved legislation campaigned against going to the polls so that, despite a swiping majority among voters, abolishers did not reach the quorum of the majority of Italians registered to vote and thus the legislation was not abolished. Likely, if the decision had been taken on the basis of the intensity of preferences, as cost-benefit analysis does, the outcome would have been different.

Whilst limiting access to ARTs services, the legislation enacted in 2004 (L. 40/2004) provided public coverage to ARTs and created a hypothecated annual funding line of € 6.8 Million aimed at "facilitating" access to fertility services. This funding is very limited, as it will cover about 3,000 IVF cycles (if a full costing method is used and costs of stimulation treatments are included). However, it signals the decision to include ARTs in the Italian basic package (Torbica and Fattore, 2005).

In the 1990s IVF was a preferred target of explicit rationing exercises. The Dutch committee on choices in health care (Dunning, 1992) used the IVF to exemplify how certain services could be excluded from basic coverage on the ground that they are not a medical necessity. The basic health plan proposed by the first Clinton Administration in the US did not include IVF and the Oregon list ranked IVF at the very bottom (Spar, 2006). In 1994 almost a quarter of English Health Authorities decided not to purchase IVF (Wiles and Patel, 1995). About ten years later the situation appears to be different. Over time IVF and other sophisticated ARTs have gained acceptance and have been increasingly funded with public money, as it shows the English NHS (NICE, 2005).

The availability of IVF services under the NHS still greatly varies across the country (NICE, 2005). IVF is a classic example of rationing by postcode: in some Health Authorities it is available and in others, it is not, without any clear rationale. In such situations, the National Institute for Health and Clinical Excellence (NICE) was asked to issue a guideline on assessment and treatment of people with fertility problems. The guideline was published in February 2004 and included indications about the use of IVF services. Basically, it suggested that up to three cycles should be offered to couples in which the woman is aged 23-39 years and who have identified cause for their fertility problems or who have infertility of at least 5 years' duration. When the guideline was released John Reid, the Secretary of State for Health of England, announced that all Primary Care Trusts that were not currently offering IVF should be in the position to offer 1 cycle to those eligible, with the longer-term aim for full implementation and 3 cycles for patients eligible for treatment.

Briefly, on the basis of clinical evidence and cost-effectiveness studies NICE suggested the inclusion of IVF in the package of services that the English NHS should offer to citizens (NICE, 2005). This policy is consistent with the results of our study in Italy. However, it does not appear that cost-benefit studies have been used to prepare the NICE guideline. In our opinion, cost-benefit analysis should be openly used to provide evidence by authorities in charge of summarising evidence to be used in policy making. Cost-benefit analysis is significantly different from ACE and ACU because it is the only valuation technique that ha no need to provide anchor points as a means of interpreting values (Fox-Rushby, 1993). In addition, there are circumstances in which cost-effectiveness and cost-utility are inappropriate, for example because the outcome of the intervention cannot be measured in terms of years of life gained or the process of going through the service may produce important welfare gains. In such cases, cost-benefit analysis is the preferred method of economic evaluation. Reproductive medicine is surely such a case and we suspect that emerging areas of medicine may pose similar problems and may require cost-benefit analysis rather than cost-effecttiveness and cost-utility analysis (e.g. genetic testing and cosmetic surgery).

In general, IVF and other Assisted Reproductive Techniques are now part of the normal coverage of statutory systems in Europe. Evidence on effectiveness and cost-effectiveness has probably favoured this. Whatever the strength of this evidence it would never suffice to give an answer to the question of whether the benefits of IVF outweigh its costs. Only cost-benefit analysis can do it. Therefore, we strongly encourage that NICE and other agencies in charge of giving direction on the content of coverage of statutory systems accept the framework of cost-benefit analysis and use sound cost-benefit studies when making recommendations.

As mentioned earlier, this study provides evidence in favour of funding IVF services to infertile couples. Our analysis shows that, following a series of reasonable assumptions, benefits greatly outweigh costs. Interestingly, however, our data clearly shows that in a real referendum a new tax sufficient to cover the costs would be approved with a very tiny majority, if any. As a consequence, while cost-benefit analysis is strongly in favour of the programme, popular voting for the programme might give a different result. Mathematically, this derives from the fact that the mean value is much larger than the median value given the large number of "0" WTP and the significant number of respondents willing to pay the maximum amount offered. Concretely, the issue concerns the impact that a limited number of respondents with a very high WTP have on results. About 600 respondents (10% of the total sample) revealed to be willing to pay € 200 or above for the programme. According to the conservative assumption that € 200 is the maximum amount that these respondents are willing to pay, about 40% of total WTP derives from the answer of these respondents. This minority of the sample determines the results of the cost-benefit analysis!

Basically, in cost-benefit analysis WTP amounts are similar to a weighing system of the voting procedure. Each respondent expresses a position that can be interpreted in terms of being in favour or against the programme and the extent of the stated WTP weights the position. To a certain extent contingent valuation can be assimilated to a procedure to measure the intensity for public choices. Indeed, WTP is a metric of preferences.

In terms of policy implications, however, aggregating individual WTP to derive a mean WTP to be compared to costs presents a major problem: the weighing system favours the opinions of the most affluent. As they have more resources available, they have higher WTP to state in contingent valuation surveys. This means that the point of view of the most affluent individuals count more than that of the less affluent. In this respect cost-benefit analysis violates the "one head –one vote" principle of modern democracy.

In our study the minority strongly in favour of the sample is also overrepresented of affluent and well educated individuals. Although sufferance due to infertility hits people across social groups and conditions, our data shows that the higher socio-economic groups are also those that are more acquainted with infertility and IVF. Likely, postponing parenthood and thus being at higher risk of infertility, is mainly an issue among highly educated professionals and middle-class couples.

The assumption underlying this research is that sound economic analysis can provide guidance to policymaking. Consistently, we used cost-benefit analysis to provide original evidence about the desirability of a publicly funded programme providing IVF to Italian infertile couples. Our data robustly indicates that such programme would imply a welfare gain for Italians. This is the main message for policy maker. Nevertheless, evidence-based policymaking cannot be based on naïve assumptions about the neutrality of theories and research methods. A major issue to be taken into consideration in cost-benefit studies based on stated or revealed preferences is that aggregate benefits measures give a louder voice to affluent individuals so creating important equity concerns.

Bibliography

Acton JP (1973). *Evaluating Public Programs to Save Lives. The case of heart attacks.* Report R95ORC. Santa Monica: Rand Corporation.

Adamson GD, de Mouzon J, Lancaster P, Nygren KG, Sullivan E, Zegers-Hochschild A (2006). World collaborative report on in vitro fertilization, 2000. *Fertility and Sterility*; 85(6): 1586-622.

Adduce A, Lorenzoni L (2004). Metodologia e primi risultati di un indagine ministeriale sui costi delle prestazioni di ricovero ospedaliero. *Politiche Sanitarie*; 5(4): 158-172.

Agresti A., Finlay B. (1984). *Statistical methods for the social sciences.* Second Edition. San Francisco: Dellen Publishing Company.

American Society for Reproductive Medicine (2000). 1998 Assisted Reproductive Technology Success rates. National Summary and Fertility Clinic Reports. http://www.cdc.gov/nccdphp/drh/art.htm. (20 October 2006).

Amin M, Khondoker F (2004). A contingent valuation study to estimate the parental willingness-to-pay for childhood diarrhoea and gender bias among rural households in India. *Health research policy and systems* 2: 3.

Anthony RN, Welsh GA, Reece JS (1985). *Fundamentals of Management Accounting.* Fourth Edition. Homewood (Ill): Irwin.

Anthony RN, Young DW (1988). *Management Control in Nonprofit Organisations.* Homewood (Ill): Irwin.

Appel LJ, Steinberg EP, Powe NR, Anderson GF, Dweyer SA, Faden RR (1990). Risk Reduction From Low Osmolality Media. What Do Patients Think It Is Worth. *Medical Care*; 28(4): 324-337.

Araña JE, León CJ. Willingness to pay for health risk reduction in the context of altruism. *Health Economics*;11(7): 623-35.

Arrow K, Solow R, Leamer E, Portney P, Randner R, Schuman H (1993). Report of the NOAA Panel on Contingent Valuation. *Federal Register*; 58(10): 4602-4614.

Asenso-Okyere WK, Osei-Akoto I, Anum A, Appiah EN (1997). Willingness to pay for health insurance in a developing economy. A pilot study of the informal sector of Ghana using contingent valuation. *Health Policy*; 42: 223-237.

Asfaw A, von Braun J (2005). Innovation in Health Care Financing: New Evidence on the Prospect of Community Health Insurance Schemes in the Rural Areas of Ethiopia. *International Journal of Health Care Finance and Economics*; 5: 241-253.

Asgary A, Willis K, Taghvaei AA, Rafeian M (2004). Estimating rural house-hold's willingness to pay for health insurance. *European Journal of Health Economics*; 5: 209-215.

Barner JC, Branvold A. Patients' willingness to pay for pharmacist-provided menopause and hormone replacement therapy consultations. *Research in social & administrative pharmacy*; 1(1): 77-100.

Beier HM, Beckman JO (1991). Implications and consequences of German Embryo Protection Act. *Human Reproduction*; 6(4):607-8.

Bennett J, Morrison M, Blamey R (1998). Testing the Validity of Responses to Contingent Valuation Questioning; *Australian Journal of Agricultural and Resource Economics*; 42(2): 131-48.

Bennett R, Tranter R (1998). The Dilemma concerning Choice of Contingent Valuation Willingness-to-Pay Elicitation Format; *Journal of Environmental Planning and Management*; 41(2): 253-57.

Berrens RP, Bohara A, Jenkins Smith H, Silva C, Weimer D (2004). Information and effort in contingent valuation surveys: application to global climate change using national internet samples. *Journal of Environmental Economics and Management*; 47: 331-63.

Berwick DM, Weinstein M C (1985). What Do Patients Value? Willingness to Pay for U1trasound in Normal Pregnancy. *Medical Care*; 23(7): 881-893.

Bhatia MR, Fox-Rushby JA. (2001). Willingness to pay for treated mosquito nets in Surat, India: the design and descriptive analysis of a household survey. *Health Policy Planning*; 17(4):402-11.

Bhatia M R, Fox-Rushby J A (2003). Validity of willingness to pay: hypothetical versus actual payment. *Applied Economics Letters*; 10: 737-740.

Birch S, Donaldosn C (2003). Valuing the benefits and costs of health care programmes: where's the "extra" in extra-welfarism? *Social Science and Medicine* 56: 1121-1133.

Bishai D, Pariyo G, Insworth A, Hill K (2004). Determinants of personal demand for an AIDS vaccine in Uganda: contingent valuation survey. *Bulletin of the World Health Organisation*; 82(9): 652-660.

Bishop RC, Heberlein TA (1978). Measuring values of extra-market goods: are indirect measures biased? *American Journal of Agricultural Economics*; 61: 936-940.

Blumenschein K, Johannesson M, Yokoyama KK, Freeman PR (2001). Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. *Journal of Health Economics*; 20: 441-457.

Bohara A, Kerkvliet J, Berrens RP (2001). Adressing Negative Willingness To Pay in Dichotomous Choice Contingent Valuation. *Environmental and Resource Economics;* 20: 173-195.

Boyle KJ, Desvousges WH, Johnson FR, Dunford RW, Hudson SP (1994). An investigation of part-whole biases in contingent valuation studies. *Journal of Environmental Economics and Management*; 27: 1330-1337.

Borisova NN; Goodman AC (2003). Measuring the value of time for methadone maintenance clients: willingness to pay, willingness to accept, and the wage rate. *Health Economics*; 12 (4): 323-34.

Bradford D W, Kleit A N, Krousel-Wood M A, Re R M (2004). Willingness to pay for telemedicine assessed by the double-bounded dichotomous choice method. *Journal of Telemedicine and Telecare*; 10: 325-330.

Brent R.J. (2003). *Cost-benefit Analysis and Health Care Evaluations*. Cheltenham (UK): Edward Elgar.

Brouwer R, Bateman IJ (2005). Benefit transfer of willingness to pay estimates and function for health-risk reductions: a cross-country study. *Journal of Health Economics*; 24: 591-611.

Brown TC (1996). Which response format reveals the truth about donations to a public good?. *Land Economics*; 72: 152–166.

Bryan S, Dolan P (2004). Discrete choice experiments in health economics. For better or for worse? *European Journal of Health Economics*; 5(3): 199-202

Cairns J (2006). Providing guidance to the NHS: The Scottish Medicines Consortium and the National Institute for Clinical Excellence compared. *Health Policy*; 76(2):134-43.

Cairns J, van der Pol M, Lloyd A (2002). Decision making heuristics and the elicitation of preferences: Being fast and frugal about the future. *Health Economics*; 11: 655-658.

Cairns J., van der Pol M (2004). Repeated follow-up as a method for reducing non-trading behaviour in discrete choice experiments. *Social Science and Medicine*; 58; 2211-2218.

Carlsson S K, Hojgard S, Lethagen S, Lindgren A, Berntorp E, Lindgren B (2004). Willingness to pay for on-demand and prophylactic treatment for severe haemophilia in Sweden. *Haemophilia*; 10: 527-541.

Cameron TA (1987). The impact of grouping coarnesses alternative grouped-data regression models. *Journal of Econometrics* 35: 37-57.

Cameron TA, Poe GL, Ethier RG, Schulze WD (2002). Alternative Non-market value-Elicitation Methods: are the Underlying Preferences the same? *Journal of Environmental Economics and Management*; 44: 391-425.

Carmines EG, Zeller R A (1979). *Realiability and Validity*. Thousands Oaks (Ca): Sage.

Carson RT, Wright J, Alberini A, Flores N (1994). A *bibliography of contingent valuation studies and papers*. In La Jolla CA (ed.) *Natural Resource Damage assessment*. Washington DC: Resource for the Future.

Centers for Disease Control and Prevention (2005). *Assisted Reproduction Technology: Success Rates 2003*. Atlanta: U.S. Department of Health and Human Services.

Chestnut LG, Keller LR, Lambert WE, Rowe RD (1996). Measuring heart patients' willingness to pay for changes in angina symptoms. *Medical Decision Making*; 16: 65-77.

Chiu L, Tang KY, Shyu WC, Chang TP (1999). The willingness of families caring for victims of stroke to pay for in-home respite care--results of a pilot study in Taiwan. *Health Policy*; 46(3):239-54.

Cho MN, Lertmaharit S, Kamol-Ratanakul P, Saul AJ (2000). Ex post and ex ante willingness to pay (WTP) for the ICT Malaria pf/pv test kit in Myanmar. *Southeast Asian Journal of Tropical Medicine and Public Health*; 31: 104-111.

Ciriacy-Wantrup SV (1947). Capital returns from soil-conservation practices. *Journal of Farm Economics*; 29, 1188-90.

Clarke PM (2002). Testing the convergent validity of the contingent valuation and travel cost methods in valuing the benefits of health care. *Health Economics*; 11(2) 117-27.

Clinch JP, Murphy A (2001). Modelling winners and losers in contingent valuation of public goods: appropriate welfare measures and econometric analysis. *The Economic Journal* 111: 420-443.

Congregazione per la Dottrina della Fede (1987). *Donum Vitae- Istruzione: il rispetto della vita umana nascente e la dignità della procreazione*. Roma: Congregazione per la Dottrina della Fede.

Cookson R (2000). Incorporating psycho-social considerations into health valuation: an experimental study. *Journal of Health Economics*; 19: 369-401.

Couper MP (2001). Web surveys: a review of issues and approaches. *Public Opinion Quarterly*; 68: 255-61.

Culyer T (1976). *Need and the National Health Service: Economics and social choice*. Oxford: Martin Robertson.

Culyer T (1991). *The normative economics of health care finance and provision.* In McGuire A, Fenn P., Mayhew K. (eds). *Providing health care: The economics of alternative systems of finance and delivery* (65-98). Oxford: Oxford University Press.

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(2): 101-18.

Davis RK (1963). Recreation planning as an economic problem. *Natural Resources Journal*; 3 : 239-249.

De Ridder A, De Graeve D (2005). Order bias in estimates of willingness to pay for drugs to treat attention-deficit/hyperactivity disorder. *European Journal of Health Economics*; 6: 146-151.

De Mouzon J, Lancaster P (1995). *World Collaborative Report 1993. 15th World Congress on Dertility and Sterility.* Montepellier: Icos.

De Vaus D, (2002). *Surveys in Social Research.* Fifth Edition. New York: Routledge.

Diener A, O'Brien B, Gafni A (1998). Health care contingent valuation studies: a review and classification of the literature. *Health Economics*; 7: 313-326.

Donaldson C (1990). Willingness to pay for publicly-provided goods: A possible Measure of Benefit. *Journal of Health Economics*; 9: 103-118.

Donaldson C (1999). Valuing the benefits of publicly-provided health care: does 'ability to pay' preclude the use of 'willingness to pay'? *Social Science and Medicine*; 49(4):551-63.

Donaldson C, Shackley P, Abdalla M, Miedzybrodzka Z (1995). Willingness to pay for antenatal screening for Cystic Fibrosis. *Health Economics*; 4: 439-452.

Donaldson C, Shackley P, Abdalla M, Miedzybrodzka Z, Semper J (1994). *The use of willingness to pay along clinical trials.* Discussion Paper 01/94. Aberdeen: Health Economics Research Unit, University of Aberdeen.

Donaldson C, Shackley P, Abdalla M (1997). Using willingness to pay to value close substitutes: carrier screening for cystic fibrosis revisited. *Health Economics*; 6(2):145-59.

Donaldson C, Hundley V, Mapp T (1998). Willingness to pay: a method for measuring preferences for maternity care? *Birth*; 25(1): 32-9.

Donaldson C, Jones AM, Mapp TJ, Olsen JA (1998). Limited dependent variables in willingness to pay studies: applications in health care. *Applied Economics*; 30: 667-677.

Dong H, Kouyate B, Cairns J, Mugisha F, Sauerborn R (2003a). Willingness-to-pay for community-based insurance in Burkina Faso. *Health Economics*; 12: 849-862.

Dong H, Kouyate B, Cairns J, Sauerborn R (2003b). A comparison of the reliability of the take-it-or-leave-it and the bidding game approaches to estimating willingness-to-pay in a rural population in Wets Africa. *Social Science and Medicine*; 56: 2181-2189.

Dong H, Kouyate B, Cairns J, Sauerborn R (2004). Differential willingness for household to pay community-based health insurance premia for themselves and other household members. *Health Policy and Planning*; 19(2): 120-126.

Dong H, Kouyate B, Cairns J, Sauerborn R. (2005). Inequality in willingness-to-pay for community-based health insurance. *Health Policy*; 72: 149-156.

Dranitsaris G (1997). A pilot study to evaluate the feasibility of using willingness to pay as a measure of value in cancer supportive care: an assessment of amifostine cytoprotection. *Supportive Care in Cancer*; 5(6): 489-99.

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 Health Technology Assessment in Health Care*;15(1):108-22.

Dranitsaris G, Longo CJ, Grossman LD (2007). The economic value of a new insulin preparation, Humalog Mix 25. Measured by a willingness-to-pay approach. *Pharmacoeconomics*;18(3):275-87.

Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Second Edition. Oxford: Oxford University Press. 1997.

Dunning AJ (1992). *Choices in Health Care*. Rijswijk (The Netherlands): Ministry of Welfare, Health and Cultural Affairs.

Eastaugh SR (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.

Eckerlund I, Johannesson M, Johansson P, Tambour M, Zethraeus K (1995). Value for money? A contingent valuation study of the optimal size of the Swedish health care budget. *Health Policy* 34: 135-143.

Edwards RG (1983). *The current clinical and ethical situation of human conception in vitro*. In Carter CD (ed.). *Developments in Human Reproduction and their Eugenic, Ethical Implications*. London: Academic Press.

Edwards RG (1998) *Introduction and development of IVF and its ethical regulation*. In Hildt E, Mieth D (eds). *In Vitro Fertilisation in the 1990s. Towards a medical, social and ethical evaluation*. Aldershot: Ashgate.

Eastaugh SR (2000). Willingness to pay in treatment of bleeding disorders. *International Journal of Technology Assessment in Health Care*;16(2):706-10.

Evans D (1995). Infertility and the NHS. Purchasers should avoid the moral high ground. *British Medical Journal*; 311: 1586.

Fasouliotis SJ, Schenker JG (1999). Social Aspects in Assisted Reproduction. *Human Reproduction Update*; 5(1): 26-39.

Fattore G (1999). *Cost-containment and reforms in the Italian National Health Service*. In: Mossialos E, Le Grand J (eds). *Health care and cost containment in the European Union*. Aldershot: Ashgate; 513-546.

Fattore G, Percudani M, Belloni D, Contini A (1997). I costi delle prestazioni in psichiatria: un'analisi condotta in un Centro Psico-Sociale (CPS) pubblico in Lombardia. *Epidemiologia e Psichiatria Sociale*; 6(2): 139-147.

Fattore G, Torbica A (2006). The inpatient reimbursement system in Italy: how do tariffs relate to costs? *Health Care Management Science*; 9: 251-258.

Fautrel B, Clarke AE, Guillemin F, Adam V, St-Pierre Y, Panaritis T, Fortin P R, Menard H A, Donaldson C, Penrod J R (2005). Valuing a hypothetical cure for rheumatoid arthritis using the contingent valuation methodology: the patient perspective. *The Journal of Rheumatology* 32: 443-453.

Finkelstein EA, Brown DS, Avidor Y, Takeuchi AH (2005). The role of price, sociodemographic factors, and health in the demand for bariatric surgery. *The American Journal of Managed Care*; 11(10): 630-637.

Finkler SA (1982). The distinction between costs and charges. *Annals of Internal Medicine*: 102-109.

Fisher M, Marks A, Trieller K, Brody R (1985). Are adolescents able and willing to pay the fee for confidential care. *Journal of Paediatrics*; 170: 480-483.

Foreit JR, Fleischman Foreit KG (2003). The reliability and validity of willingness to pay surveys for reproductive health pricing decisions in developing countries. *Health Policy*; 63: 37-47.

Forsythe S, Arthur G, Ngatia G, Mutemi R, Odhiambo J, Gilks C (2002). Assessing the cost and willingness to pay for voluntary HIV counselling and testing in Kenya. *Health Policy and Plannning*; 17(2): 187-95.

Fox-Rushby J (1993). Appraising the use of contingent valuation: a note in response. *Health Economics* 2; 361-362.

France G, Taroni F, Donatini A. The Italian health-care system. *Health Economics* 2005; 14(S1): S187-202.

Frew EJ, Wolstenholme JL, Whynes DK. (2004). Comparing willingness-to-pay: bidding game format versus open-ended and payment scale formats. *Health Policy*; 68(3):289-98

Gafni A (1991). Willingness-to-pay as a measure of benefit. Relevant questions in the context of public decision making about health care programs. *Medical Care*; 29(12), 1246-1252.

Gafni A (1998). Willingness to pay. What's in a name? *Pharmacoeconomics*; 14 (5): 465-470.

Giacomini M., Hurley J., Stoddart G (2000). The many meanings of deinsuring a health service: the case of in vitro fertilization in Ontario. *Social Science and Medicine*; (50)10: 1485-500.

Golan EH, Shechter M (1993). Contingent valuation of supplemental health care in Israel. *Medical Decision Making* 13: 302-310.

Gold M R, Siegel J E, Russel L B, Weinstein M C (1996). *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.

Granberg M, Wikland M, Nilsson L, Hamberger L (1995). Couples' willingness to pay for IVF. *Acta Obstetrica et Gynaecologica Scandinavica* 74: 199-202.

Graves N, Walker D, Rainer N, Hutchings A, Roberts JA (2002). Cost data for individual patients included in clinical studies: no amount of statistical analysis can compensate for inadequate costing methods. *Health Economics*; 11(8): 735-9.

Greenberg D, Bakhai A, Neumann PJ, Cohen DJ (2004). Willingness to pay for avoiding coronary restenosis and repeat revascularization: results from a contingent valuation study. *Health Policy*; 70: 207-216.

Grimes DS (1988). Value of a negative cervical smear. *British Medical Journal*; 296: 1363.

Gyldmark M, Morrison GC (2001). Demand for health care in Denmark: results for a national sample survey using contingent valuation. *Social Science and Medicine*; 53: 1023-1036.

Haan (1991). Effects and costs of in-vitro fertilization. *International Journal of Technology Assessment in Health Care*; 7(4): 585-593.

Hackl F, Pruckner G J (2005). Warm glow, free-riding and vehicle neutrality in health-related contingent valuation study. *Health Economics* 14: 293-306.

Hakim Z., Pathak D.S. (1999). Modelling the Euroquol data: a comparison of discrete choice conjoint and conditional preference modelling. *Health Economics*; 1999; 8(2): 103-116.

Ham C, Robert G (2003). *Conclusions*. In Ham C, Robert G (eds). *Reasonable rationing: International experience of priority setting in health care*. Maidenhead: Open University Press.

Hamelsky SW, Lipton RB, Stewart WF (2005). An assessment of the burden of migraine using the willingness to pay model. *Cephalalgia*; 25: 87-100.

Hammerschmidt T, Zeitler HP, Leidl R (2004). A utility-theoretic approach to the aggregation of willingness to pay measured in decomposed scenarios: development and empirical test. *Health Economics*; 13: 345-361.

Hammit JK, Graham JD (1999). Willingness to pay for health protection: inadequate sensitivity to probability. *Journal of Risk and Uncertainty*; 8: 33-62.

Hanemann WM (1996). *Theory versus data in the contingent valuation debate*, in Bjornstad DJ, Kahn JR (Eds) *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs*. Brookfield (MA): Elgar.

Hanemann WM, Kristrom B (1995). *Preference uncertainty, optimal design and spikes*. In Johansson P, Kristrom B, Maler K. *Current issues in environmental economics*. Manchester: Manchetser University Press.

Hanley N, Ryan M, Wright R (2003). Estimating the monetary value of health care: lessons from environmental economics. *Health Economics*; 12: 3-16.

HFEA (Human Fertility and Embryology Authority) (2000). Ninth Annual Report and Account. (http://www.hfea.gov.uk/) (23 August 2002).

HFEA (Human Fertility and Embryology Authority). Facing the future. Annual report 2004/05. 2005. (http://www.hfea.gov.uk/) (16 September 2006).

HFEA (Human Fertility and Embryology Authority). Facts and Figures 2003-04. 2006. (http://www.hfea.gov.uk/cps/rde/xbcr/SID-3F57D79B-4A23A020/hfea/facts_and_figures.pdf) (16 September 2006).

Ho JJ, Liu JT, Wang JD (2005). Stated preferences for the removal of physical pain resulting from permanently disabling occupational injuries. A contingent valuation study of Taiwan. *Accident Analysis and Prevention*; 37: 537-548.

Holmes TP, Kramer RA (1995). An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation. *Journal of Environmental Economics and Management*; 29: 121-132.

Horgren CT, Sundem GL, Stratton WO (1996). *Introduction to Management Accounting*. Tenth Edition. London: Prentice Hall.

Huang JC, Smith VK (1998) Monte Carlo benchmarks for discrete response valuation methods. *Land Economics* 79: 186–202.

Holmes TP, Kramer RA (1995). An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation. *Journal of Environmental. Economics and Management*; 29: 121–132.

Human Fertilisation and Embriology Act (1990), *Chapter 37*. London: Her Majesty's Stationery Office.

Iglesias T (1990). *IVF and Justice. Moral, social and legal issues related to human In Vitro Fertilisation*. London: The Linacre Center.

ISTAT (2005). Compendio Statistico 2004. Roma: Poligrafico dello Stato.

ISTAT (2007). La vita quotidiana nel 2006. Roma: Poligrafico dello Stato.

Jacobs P, Bachynsky J. Costing methods in the Canadian literature on the economic evaluation of health care. A survey and assessment. *International Journal of Technology Assessment in Health Care*: 1996; 12(4): 721-734.

Johannesson M (1996a). *Theory and methods of economic evaluation of health care*. Dordrecht: Kluwer.

Johannesson M (1996b). A note on the relationship between ex ante and expected willingness to pay for health care. *Social Science and Medicine*; 42(3), 305-311.

Johannesson M, Fagerberg B (1992). A health-economic comparison of diet and drug treatment in obese men with mild hypertension. *Journal of Hypertension*; 10: 1063-1070.

Johannesson M, Johansson P, Kristrom B, Borgquist L, Jonsson B (1993b). Willingness to pay for lipid lowering: a health production function approach. *Applied Economics*; 25: 1023-1031.

Johannesson M, Johansson P, Kristrom B, Gerdtham U (1993a). Willingness to pay for hypertensive therapy - Further results. *Journal of Health Economics*; 12: 95-108.

Johannesson M, Jonsson B (1991). Economic evaluation in health care: is there a role for cost-benefit analysis? *Health Policy*; 17: 1-23.

Johannesson M, Jonsson B, Borgquist L (1991). Willingness to pay for antihypertensive therapy - Results of a Swedish pilot study. *Journal of Health Economics*; 10: 461-474.

Jones-Lee M W, Hammerton M, Philips P R (1995). The value of safety: Results of a national sample survey. *The Economic Journal*; 95: 49-72.

Kahneman D, Knetsch JL (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of Environment Economics and Management*; 22: 57-70.

Kanninen BJ (1995). Bias in discrete response contingent valuation. *Journal of Environmental. Economics and. Management*; 28: 114–125

Kartman B, Andersson F, Johannesson M (1996a). Willingness to pay for reduction in angina pectoris attacks, *Medical Decision Making*; 16: 248-253.

Kartman B, Stalhammar N, Johannesson M (1996b). Valuation of health changes with the contingent valuation method: a test of scope and question order effects. *Health Economics*; 5: 531-541.

Klein R, Day P, Redmayne S (1996). *Managing scarcity: Priority setting and rationing in the National Health Service*. Buckingham: Open University Press.

Klose T (1999). The contingent valuation method in health care. *Health Policy*; 47(2): 97-123.

Koch L (1998). *Two decades of IVF: A critical appraisal.* In Hildt E, Mieth D (eds). *In Vitro Fertilisation in the 1990s. Towards a medical, social and ethical evaluation.* Aldershot: Ashgate.

Kramer RA, Mercer DE (1997). Valuing a global environmental good: U.S. residents' willingness to pay to protect tropical rain forest. *Land Economics*; 73 (2): 196-210.

Kristrom B (1997). Spike Models in Contingent Valuation. *American Journal of Agricultural Economics*; 79: 1013-1023.

Lancsar E, Donaldson C (2004). Discrete choice experiments in health economics: distinguishing between the method and its application; *European Journal of Health Economics* 5 (3): 199-202.

Lansac J (1996). French law concerning medically-assisted reproduction. *Human Reproduction*; 11(9):1843-7.

Le Grand J (2003), *Motivation, agency, and Public Policy. Of knights & knaves, pawns & queens.* Oxford: Oxford University Press.

Le Grand J, Bartlett W. (1993), *Quasi –markets and social policy: The way forward?* In Le Grand J, Bartlett W (der). *Quasi-markets and social policy.* London: Mac Millan.

Le Grand J, Propper C, Robinson R (1992). *The Economics of Social Problems.* London: MacMillan.

Lee GM, Salomon JA, Le Baron CW, Lieu TA (2005). Health-state valuations for pertussis : methods for valuing short-term health states. *Health and Quality of Life Outcomes*; 3: 17.

Lee SJ, Neumann PJ, Curchill WH, Cannon ME, Weinstein MC, Johannesson M (1996). Patients' willingness to pay for autologous blood donation. *Health Policy*; 40: 1-12.

Lindholm L, Rosen M, Hellsten G (1994). Are people willing to pay for a community-based preventive program. *International Journal of Technology Assessment in Health Care*; 10(2): 317-324.

Li H, Berrens RP, Bohara AK, Hank C, et al. (2004). Would developing country commitments affect US households' support for a modified Kyoto Protocol? *Ecological Economics*; Volume 48: 329-343.

Liu JT, Hammitt JK, Wang JD, Liu JL (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.

Liu JT, Hammitt JK, Wang JD, Tsou MW (2005). Valuation of the risk of SARS in Taiwan. *Health Economics*; 14: 83-91.

Long J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables.* Thousands Oak (Ca): Sage.

Luchini S, Protière C, Moatti JP (2003). Eliciting several willingness to pay in a single contingent valuation survey: application to health care. *Health Economics*; 12: 51-64.

Luce BR, Elixhauser A (1990). Estimating costs in the economic evaluation of medical technologies. *International Journal of Technology Assessment in Health Care*; 6: 57-75.

Maddala GS (1983). *Limited-dependent and qualitative variables in econometrics.* New York: Cambridge University Press.

Mantovani LG, Belisari A, Szucs TD (1999). Pharmaco-economics aspects of in-vitro fertilization in Italy. *Human Reproduction*; 14 (4): 953-958.

Marta-Pedroso C, Freitas H, Domingos T (2007). Testing for the Survey Mode Effect on Contingent Valuation Data Quality: A Case Study of Web Based versus In-Person Interviews. *Ecological Economics*; 62: 388-98.

Masiye F, Rehnberg C (2005). The economic value of an improved malaria treatment programme in Zambia: results from a contingent valuation survey. *Malaria Journal*; 4:60.

Mataria A, Donaldson C, Luchini S, Moatti JP (2004). A stated preference approach to assessing health care-quality improvements in Palestine: from theoretical validity to policy implications. *Journal of Health Economics*; 23: 1285-1311.

Mathiyazhagan K (1998). Willingness to pay for rural health insurance through community participation in India. *International Journal of Health Planning And Management*; 13 (1): 47-67.

Matthews DC, Birch S, Gafni A, Di Censo A. Willingness to pay for periodontal therapy: development and testing of an instrument. *Journal of Public Health Dentistry;* 59(1): 44-51.

McIntosh E, Donaldson C, Ryan M (1999). Recent advances in the methods of cost-benefit analysis in healthcare. *Pharmacoeconomics* 15 (4): 357-367.

McKenzie L, Cairns J, Osman L (2001). Symptom-based outcome measures for asthma: the use of discrete choice methods to assess patient preferences. *Health Policy*; 57: 193-204.

Mechanic D (1995). Dilemmas in rationing health care services: the case for implicit rationing. *British Medical Journal*; 310: 1655-99.

Medical Research International (1999). In vitro fertilization/embryo transfer in the United States: 1987 results from the National IVF/ET Registry. *Fertility and Sterility*; 51: 13-18.

Miedzybroddzka 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.

Mills A, Fox-Rushby J, Aikins M, D'Alessandro U, Cham K, Greenwood B (1994). Financing mechanisms for village activities in the Gambia and their implications for financing insecticide for bednet impregnation. *Journal of Tropical Medicine and Hygiene*; 97: 325-332.

Mishan E J (1971). *Cost-benefit analysis.* New York: Praeger.

Mitchell RC, Carson RT (1981). *An experiment in determining willingness to pay for national water quality improvements: Draft report to the US Environmental Protection Agency.* Washington: Environmental Protection Agency.

Mitchell RC, Carson RT (1989). *Using surveys to value public goods. The contingent valuation method.* Washington: Resources for the Future.

Morris J; Hammitt JK (2001). Using life expectancy to communicate benefits of health care programs in contingent valuation studies. *Medical Decision Making*; 21(6): 468-78

National Collaborating Centres for Women's and Children's Health (2004), *Fertility assessment and treatment for people with fertility problems.* London: ROCG Press.

Neumann PJ, Gharib SD, Weinstein M C (1994). The cost of a successful delivery with In Vitro Fertilisation. *New England Journal of Medicine*; 331: 239-243.

Neumann PJ, Johannesson M (1994). The willingness to Pay for In Vitro Fertilisation: A Pilot Study Using Contingent Valuation. *Medical Care*; 32(7): 686-699.

New B, Le Grand J (1996). *Rationing in the NHS. Principles and pragmatism.* London: King's Fund.

NHSME, Costing for Contracting. EL(93)26. Leeds. 1993.

NICE (National Institute for Clinical Excellence) (2005). *Clinical Guideline 11: Fertility.* www.nice.org.uk/CG011 (5 August 2006).

Nocera S, Bonato D, Telser H (2002). The contingency of contingent valuation. How much are people willing to pay against Alzheimer's disease? *International Journal of Health Care Finance and Economics*; 2(3): 219-40.

Nyandieka LN, Bowden A, Wanjau J, Fox-Rushby J. (2002). Managing a household survey: a practical example from the KENWOL survey. *Health Policy and Planning*; 17(2): 207-212.

O'Brien B, Novosel S, Torrance G, Streiner D (1995). Assessing the economic value of a new antidepressant. A willingness-to-pay approach. *Pharmacoeconomics*; 8(1): 34-45.

O'Brien B, Gafni A (1996). When do the "dollars" make sense? Toward a conceptual framework for contingent valuation studies in health care. *Medical Decision Making* 16(3), 288-299

O'Brien BJ, Goeree R, Gafni A, et al. Assessing the value of a new pharmaceutical. A feasibility study of contingent valuation in managed care. *Medical C*are; 36(3): 370-84.

O'Conor RM; Blomquist GC (1997). Measurement of consumer-patient preferences using a hybrid contingent valuation method. *Journal of Health Economics*; 16(6): 667-83.

O'Brien B, Viramontes JL (1994). Willingness to Pay: A Valid and Reliable Measure of Health Preference? *Medical Decision Making*; 14: 289-297.

Olsen JA (1997). Aiding priority setting in health care: is there a role for the contingent valuation method? *Health Economics*; 6: 603-612.

Olsen JA, Donaldson C (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.

Olsen JA, Kidholm K, Donaldson C, Shackley P, Olsen JA (2004a). Willingness to pay for public health care: a comparison of two approaches. *Health Policy*;70(2):217-28.

Olsen JA, Donaldson C, Pereira J (2004b). The insensitivity of "willingness-to-pay" to the size of the good: New evidence for health care. *Journal of Economic Psychology 25*: 445-460.

Olsen J A, Donaldson C, Shackley P (2005). Implicit versus explicit ranking: on inferring ordinal preferences for health care programmes based on differences in willingness-to-pay. *Journal of Health Economics*; 24: 990-996.

Olsen J A, Kidholm K, Donaldson C, Shackley P (2004). Willingness to pay for public health care: a comparison of two approaches. *Health Policy*; 70: 217-228.

Olsen JA, Smith RD (2001). Theory versus practice: a review of willingness to pay in health and health care. *Health Economics*; 10: 39-52.

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.

Onwujekwe O (2004). Criterion and content validity of a novel structured haggling contingent valuation question format versus the bidding game and binary with follow-up format. *Social Science and Medicine*; 58: 525-537.

Onwujekwe O, Chima R, Shu E, Nwagbo D, Okonkwo P (1998). Hypothetical and actual willingness to pay for insecticide-treated nets in five Nigerian communities. *Tropical Medicine & International Health*; 6(7): 545-53.

Onwujekwe O, Chima R, Shu E, Nwagbo D, Okonkwo P (2001). Hypothetical and actual willingness to pay for insecticide-treated nets in five Nigerian communities. *Tropical Medicine & International Health*; 6(7): 545-53.

Onwujekwe O, Chima R, Shu E, Nwagbo D, Akpala C, Okonkwo P (2002). Altruistic willingness to pay in community-based sales of insecticide-treated nets exists in Nigeria. *Social Science & Medicine*; 54: 519-527.

Onwujekwe O, Fox-Rushby J, Hanson K (2004). Valuing the benefits of a health intervention using three different approaches to contingent valuation: re-treatment of mosquito bed-nets in Nigeria. *Journal of Health Services Research & Policy* 9 (2): 67-75.

Onwujekwe O, Fox-Rushby J, Hanson K (2005a). Inter-rater and test-retest reliability of three contingent valuation question formats in south-east Nigeria. *Health Economics*; 14: 529-536.

Onwujekwe O, Fox-Rushby J, Hanson K (2008). Construct validity of the bidding game, binary with follow-up, and a novel structured haggling question format in determining willingness to pay for insecticide-treated mosquito nets. *Medical Decision Making*; 28(1): 90-101.

Onwujekwe O, Hanson K, Fox-Rushby JA. Who buys insecticide-treated nets? Implications for increasing coverage in Nigeria. *Health Policy and Planning*. 2003;18(3):279-89.

Onwujekwe O, Hanson K, Fox-Rushby J (2005b). Do divergences between stated and actual willingness to pay signify the existence of bias in contingent valuation surveys? *Social Science & Medicine*; 60: 525-536.

Onwujekwe O, Nwagbo D (2002). Investing starting-point bias: a survey of willingness to pay for insecticide-treated nets. *Social Science & Medicine*; 55: 2121-2130.

Onwujekwe O, Shu EN, Okonkwo PO (1999). Willingness to pay for the maintenance of equity in a local ivermectin distribution scheme in Toro, Northern Nigeria. *Public Health*; 113 (4): 193-4.

Onwujekwe O; Shu E; Chima R; Onyido A; Okonkwo P (2000). Willingness to pay for the retreatment of mosquito nets with insecticide in four communities of south-eastern Nigeria. *Tropical Medicine & International Health*; 5(5): 370-6.

Onwujekwe O, Uzochukwu B (2004). Stated and actual altruistic willingness to pay for insecticide-treated nets in Nigeria: validity of open-ended and binary with follow-up questions. *Health Economics*; 13: 477-492.

Ortega A, Dranitsaris G, Puodziunas AL. What are cancer patients willing to pay for prophylactic epoetin alfa? A cost-benefit analysis. *Cancer*; 83(12): 2588-96.

Osmond MH, Klassen TP, Quinn JV (1995). Economic comparison of a tissue adhesive and suturing the repair of paediatric facial lacerations. *Journal of Paediatrics*; 126(6): 892-895.

Papatheofanis FJ (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 MV (1995). *Valuing health care benefits in money terms.* In Sloan FA (ed.). *Valuing health care.* Cambridge (UK): Cambridge University Press.

Pavlova M, Groot W, Van Merode G (2004). Willingness and ability of Bulgarian consumers to pay for improved public health care services. *Applied Economics*; 36: 1117-1130.

Pennie RA, O'Connor AM, Garvock MJ, Drake ER (1991). Factors influencing the Acceptance of Hepatitis B Vaccine by Students in Health Disciplines in Ottawa. *Canadian Journal of Public Health*; 82: 12-15.

Portney PR (1994). The contingent valuation debate: why economists should care. *Journal of Economic Perspectives*; 8: 3-18.

Propper C (1990). Contingent valuation of time spent on NHS waiting lists. *The Economic Journal*; 90: 193-199.

Protière C, Donaldson C, Luchini S, Moatti J P, Shackley P (2004). The impact of information on non-health attributes on willingness to pay for multiple health care programmes. *Social Science & Medicine*; 58: 1257-1269.

Ready RC, Buzby JC, Hu D (1996). Differences between continuous and discrete contingent value estimates. *Land Economics*; 72 (3): 397-411.

Redmayne S, Klein R (1993). Rationing in practice: the case of In-Vitro Fertilisation. *British Medical Journal*; 306: 1521-1524.

Reis J, Sherman S, Macon J, Friedman B (1990). Care for the uninsured: Who should pay. *Journal of Nursing Administration*; 20(3): 16-20.

Rheingans RD, Haddix AC, Messonnier ML, Meltzer M, Mayard G, Addiss DG (2004). Willingness to pay for prevention and treatment of lymphatic filariasis in Leogane, Haiti. *Filaria Journal*; 3:2.

Robinson R (1993). Cost-benefit analysis. *British Medical Journal* 307: 924-926.

Ross CK, Steward CA, Sinacore JM (1995). A Comparative study of seven measures of patient satisfaction. *Medical Care*; 33(4): 392-406.

Ryan M (1996). Using willingness to pay to assess the benefits of assisted reproductive techniques. *Health Economics*; 5:543-558.

Ryan M (1997). Should government fund assisted reproductive techniques? A study using willingness to pay. *Applied Economics*; 29: 841-849.

Ryan M (1998), Valuing psychological factors in the provision of assisted reproductive techniques using the economic instrument of willingness to pay. *Journal of Economic Psychology*; 19: 179-204.

Ryan M (1999). Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. *Social Science and Medicine*; 48: 535-546.

Ryan M (2004). A comparison of stated preference methods for estimating monetary values. *Health Economics*; 23: 291-296.

Ryan M, Amaya-Amaya M (2005). "Threats" to and hopes for estimating benefits. *Health Economics*; 14: 609-619.

Ryan M, Ratcliffe J, Tucker J (1997). Using willingness to pay to value alternative models of antenatal care. *Social Science and Medicine*; 44(3): 371-380.

Ryan M, Scott D A, Donaldson C (2004). Valuing health care using willingness to pay: a comparison of the payment card and dichotomous choice methods. *Journal of Health Economics*; 23: 237-258.

Ryan M, Scott D, Reeves C, Bate A, van Tejilingen ER, Russell EM, Napper M, Robb CM (2001). Eliciting public preferences for health care: a systematic review of techniques. *Health Technology Assessment*; 5(5).

Ryan M, Donaldson C (1996). Assessing the costs of assisted reproductive techniques. *British Journal of Obstetrics and Gynaecology*; 103: 198-201.

Sadri H, MacKeigan LD, Leiter LA, Einarson TR (2005). Willingness to pay for inhaled insulin. A contingent valuation approach. *Pharmacoeconomics* 23(12): 1215-1227.

Schreyogg J., Stargardt T, Velasco-Garrido M, Busse R (2005). Defining the Health Benefit Basket" in nine European countries: Evidence from the European Union Health BASKET Project. *European Journal of Health Economics*; 6(S1): 2-10.

Schreyogg J, Stargardt T, Tieman O, Busse R (2006). Methods to determine reimbursement rates for diagnosis related groups (DRG): a comparison of nine European countries. *Health Care Management Science*; 9 (3):215-23.

Serour GI, Rhodes CA, Sattar MA, Aboulgar MA, Mansour RT (1999). Complications of Assisted Reproductive Techniques: A Review. *Assisted Reproduction*; 9(4): 214-232.

Shackley P, Donaldson C (2000). Willingness to pay for publicly-financed health care: how should we use the numbers? *Applied Economics*; 32: 2015-2021.

Shechter M, Kim M (1991). Valuation of Pollution Abatement Benefits: Direct and Indirect Measurement. *Journal of Urban Economics*; 30: 133-151.

Shenker JG (1997). Assisted reproduction practice in Europe: legal and ethical aspects. *Human Reproduction Update*; 3(2): 173-184.

Shenker JG, Shushan A (1996). Ethical and legal aspects of assisted reproduction practice in Asia. *Human Reproduction*; 11(4): 908-911.

Shiell A, Gold L (2003). If the price is right: vagueness and values clarification in contingent valuation. *Health Economics*; 12: 909-919.

Shiell A, Rush B (2003). Can willingness to pay capture the value of altruism? An exploration of Sen's notion of commitment. *Journal of Socio-Economics*; 32: 647-660.

Singleton RA Jr., Straits BC (1999). *Approaches to Social Research: Third Edition.* New York: Oxford University Press.

Slothuus U, Larsen ML, Junker P. 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 RD (2003). Construction of the contingent valuation market in health care: a critical assessment. *Health Economics*; 12: 609-628.

Smith RD (2005). Sensitivity to scale in contingent valuation: the importance of budget constraint. *Journal of Health Economics*; 24: 515-529.

Smith RD (2006). It's not just *what* you do, it's the *way* that you do it: the effect of different payment card formats and survey administration on willingness to pay for health gain. *Health Economics*; 15 (3): 281-93.

Smith VK (2006). *Fifty years of contingent valuation.* In Alberini A and Khan JR (edr). *Handbook on Contingent Valuation.* Cheltenham: Edwar Elgar.

Sorum PC (1999). Measuring patient preferences by willingness to pay to avoid: the case of acute otitis media. *Medical Decision Making*;19 (1):27-37.

Spar D (2006). *The Baby Business. Elite eggs, designer genes, and the thriving commerce of the conception*, Boston: Harvard Business School Press.

Stalhammar N (1996). An empirical note on willingness to pay and starting-point Bias. *Medical Decision Making*; 16: 242-247.

Stewart JM, O'Shea E, Donaldson C. (2002). Do ordering effects matter in willingness-to-pay studies in healthcare? *Journal of Health Economics*; 21: 585-599.

Stiglitz JE. *Economics of the public sector.* Third Edition. New York: Norton. 1988.

Suh DC (2000). Consumers' willingness to pay for pharmacy services that reduce risk of medication-related problems. *Journal of the American Pharmaceutical Association*; 40(6): 818-27.

Suraratdecha C, Ainsworth M, Tangacharoensathien V, Whittington D (2005). The private demand for AIDS vaccine in Thailand. *Health Policy*;71: 271-287.

Takemura S, Ohida T, Sone T, Fukada T, Takemura Y (2005). Influences of absence of random assignment of bids on estimating willingness to pay using a discrete-choice question. *Health Economics*; 14: 209-213.

Tambour M, Zethraeus N (1998). Nonparametric willingness-to-pay measures and confidence statements. *Medical Decision Making*; 18(3): 330-6.

Tarasiuk A, Simon T, Regev U, Reuveni H (2003). Willingness to pay for polysomnography in children with obstructive sleep apnea syndrome: a cost-benefit analysis. *Sleep* 2003; 26 (8): 1016-21.

Taylor SJ, Armour CL (2002). Acceptability of willingness to pay techniques to consumers. *Health Expectations*;5(4):341-56.

Taylor RS, Drummond MF, Salked G, Sullivan SD (2004). Inclusion of cost effectiveness in licensing requirements of new drugs: the fourth hurdle. *British medical Journal*; 329: 972-5.

Thomas R, Donaldson C, Torgerson D (2000). Who answers 'willingness to pay' questions? *Journal of Health Service Research Policy*;5(1):7-11.

Thompson MS (1986). Willingness to pay and accept risks to cure chronic disease. *American Journal of Public Health*; 76(4): 392-396.

Thompson MS, Read JL, Liang M (1984). Feasibility of willingness-to-pay measurement in chronic arthritis. *Medical Decision Making*; 4: 195-215.

Thurston HW (2006). *Non-market valuation on the internet.* In Alberini A and Khan JR (edr). *Handbook of Contingent Valuation.* Cheltenham: Edwar Elgar.

Torbica A, Fattore G (2005), "The "Essential Levels of Care" in Italy: when being explicit serves the devolution of powers", *European Journal of Health Economics*; 6: 46-52.

Torrance G, Walker V, Grossman R, Mukherjee J, Vaughan D, La Forge J, Lampron N. Economic evaluation of ciprofloxacin compared with usual antibacterial care for the treatment of acute exacerbations of chronic bronchitis in patients followed for 1 year. *Pharmacoeconomics*; 16(5 Pt 1): 499-520.

Udoff LC, Adashy EY (1999). Infertility. In Knobil E, Neill JD (eds). Encyclopedia of Rep*roduction*. San Diego: Academic Press.

Van de Ven W (1995). *Choices in health care: a contribution from the Netherlands.* In Maxwell (ed), *Rationing Health Care.* London: Churchill Livingston.

Van der Pol M., Cairns J. (1998). Establishing patient preferences for blood transfusion support: an application of conjoint analysis. *Journal of Health Services Research and Policy*; 3: 70-76.

Van der Pol M., Cairns J. (2001). Estimatine time preferences for health using discrete choice experiments. *Social Science and Medicine*; 52: 1459-1470.

Vick S, Scott A (1998). What makes a perfect agent? A pilot study of patients' preferences in the doctor-patient relationship. *Journal of Health Economics*; 17(5): 587-605.

Villari P, Fattore G, Siegel JE, Paltiel DE, Weinstein MC (1996). Economic Evaluation of HIV Testing Among Intravenous Drug Users: Analytic Framework and Application to Italy. *International Journal of Technology Assessment in Health Care*; 12(2): 329-350.

Viscusi WK (1993). The value of risks to life and death. *Journal of Economic Literature*; 31: 1912-1946.

Viscusi WK, Evans WN (1990). Utility functions that depend on health status: estimates and economic implications. *American Economic Review*; 80: 353-360.

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 and Medicine*; 43(6): 985-998.

Wagner TH, Hu TW, Dueñas GV, Pasick RJ (2000). Willingness to pay for mammography: item development and testing among five ethnic groups. *Health Policy*; 53(2):105-21.

Wagner TH, Hu T, Dueñas GV, Kaplan CP, Nguyen BH, Pasick RJ (2001). Does willingness to pay vary by race/ethnicity? An analysis using mammography among low-income women. *Health Policy*; 58(3):275-88.

Weisbrod BA (1964). Collective consumption services of individual consumption goods. *Quarterly Journal of Economics*; 878: 471-477.

Welsh MP, Poe GL (1998). Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach. *Journal of Environmental Economics and Management*; 36(2): 170-85.

Whynes DK, Frew E, Wolstenholme JL (2003). A comparison of two methods for eliciting contingent valuations of colorectal cancer screening. *Journal of Health Economics*; 22: 555-574.

Whynes DK, Philips Z, Frew E (2005). Think of a number...any number? *Health Economics*; 14: 1191-1195.

Whynes DK, Wolstenholme JL, Frew E (2004). Evidence of range bias in contingent valuation payment scales. *Health Economics*; 13: 183-190.

Wiles R., Patel H. (1995). *Report of the third national survey of NHS funding fertility services*. London: Royal College of Health.

Williams A (1983). *The economic role of "health indicators"*. In Teeling Smith (ed.). *Measuring the Social benefits of Medicine*. London: Office of Health Economics.

Whitehead JC (2006). *A practitioner's primer on the contingent valuation method*. In Alberini A and Khan JR (edr). *Fifty years of economic valuation*. Cheltenham: Edwar Elgar

Whittington D, Matsui-Santana O, Freiberger JJ, Van Houtven G, Pattanayak (2002). Private demand for a HIV/AIDS vaccine: evidence from Guadalajara, Mexico. *Vaccine*; 20(19-20): 2585-91

Wordsworth S, Ludbrook A, Caskey F, MacLeod A (2005). Collecting unit cost data in multicentre studies. Creating comparable methods. *European Journal of Health Economics*; 6 (1): 38-44.

Yeung RY, Smith RD, McGhee SM (2003). Willingness to pay and size of health benefit: an integrated model to test for 'sensitivity to scale'. *Health Economics* 12(9):791-6.

Yoo S, Kwak S (2002). Using spike model to deal with zero response data from double bounded dichotomous choice contingent valuation surveys; *Applied Economics Letters* 9: 929-932.

Zarkin GA; Cates SC; Bala M (2000). Estimating the willingness to pay for drug abuse treatment: a pilot study. *Journal of Substance Abuse Treatment*; 18 (2): 149-59.

Zethraeus N, Johannesson M, Henriksson P, Strand RT (1997). The impact of hormone replacement therapy on quality of life and willingness to pay. *British Journal of Obstetrics and Gynaecology* ; 104(10): 1191-5.

Zillich AJ; Blumenschein K; Johannesson M; Freeman P (2002). *Pharmacoeconomics*; 20(4): 257-65

Appendix A

Questionnaire and coding
(Original answers version)

DOMANDA 1: Lei conosce delle coppie che desiderano avere dei bambini e non riescono ad averli?

V71

```
        Print Format: F1
        Write Format: F1

        Value    Label

            1    si'
            2    no
```

DOMANDA 2: Lei era a conoscenza della Fertilizzazione in Vitro?

V72

```
        Print Format: F1
        Write Format: F1

        Value    Label

            1    e' la prima volta che ne sento parlare
            2    si', ne ho sentito parlare e\o letto sui giornali
            3    si', ne ho sentito parlare da amici\conoscenti\parenti
            4    si', mi sono informato\a attraverso letture specializzate
            5    si', mi sono informato\a dal medico
```

DOMANDA 3: Lei direbbe che le sue conoscenze sulla Fertilizzazione sono: (esclusi coloro che hanno risposto 'è la prima volta che ne sento parlare', codice 1 a dom. 2)

V73

```
        Print Format: F1
        Write Format: F1

        Value    Label

            1    molto scarse
            2    abbastanza scarse
            3    sufficienti
            4    abbastanza approfondite
            5    molto approfondite
```

DOMANDA 4: Lei conosce qualcuno che abbia utilizzato la Fertilizzazione in Vitro o altre terapie finalizzate a permettere a coppie non fertili di avere bambini?

V74

```
        Print Format: F1
```

Write Format: F1

Value   Label

1   si'
2   No

DOMANDA 5: Immagini di essere in una situazione in cui è sposato/a da qualche anno e in cui, malgrado desiderandoli, non riesca ad avere dei bambini. Immagini che le venga indicata la Fertilizzazione in Vitro e che le venga prospettata una probabilità del 30% di riuscire ad avere un bambino. Lei personalmente proverebbe la fertilizzazione in Vitro:

V75

Print Format: F1
Write Format: F1

Value   Label

0   Non so
1   Sicuramente no
2   Probabilmente no
3   Probabilmente si'
4   Sicuramente si'

DOMANDA 6: E, Lei proverebbe la Fertilizzazione in vitro se…
Per ciascuna delle seguenti alternative, indichi, per favore, se sì o se no, tenendo presente che:

- se fosse gratuita…

V76.1

Print Format: F1
Write Format: F1

Value   Label

1   si', proverei la Fertilizzazione in Vitro
2   no, non proverei la fertilizzazione in Vitro

- se le fosse chiesto di pagare 1 milione…

V76.2

Print Format: F1
Write Format: F1

Value   Label

1   si', proverei la Fertilizzazione in Vitro
2   no, non proverei la fertilizzazione in Vitro

- se le fosse chiesto di pagare 5 milioni…

V76.3

Print Format: F1
Write Format: F1

Value   Label

1   si', proverei la Fertilizzazione in Vitro
2   no, non proverei la fertilizzazione in Vitro

- se le fosse chiesto di pagare 10 milioni...

V76.4

Print Format: F1
Write Format: F1

Value	Label
1	si', proverei la Fertilizzazione in Vitro
2	no, non proverei la fertilizzazione in Vitro

- se le fosse chiesto di pagare 20 milioni...

V76.5

Print Format: F1
Write Format: F1

Value	Label
1	si', proverei la Fertilizzazione in Vitro
2	no, non proverei la fertilizzazione in Vitro

- se le fosse chiesto di pagare 50 milioni...

V76.6

Print Format: F1
Write Format: F1

Value	Label
1	si', proverei la Fertilizzazione in Vitro
2	no, non proverei la fertilizzazione in Vitro

SPLIT 4 CAMPIONI

DOMANDA 7: Immagini ora che in Italia si decida di fare un referendum per decidere se finanziare la Fertilizzazione in Vitro con soldi pubblici. Ogni anno il finanziamento sarebbe disponibile per circa 10.000 coppie alle quali questa terapia viene indicata. Di queste coppie circa 3.000 riuscirebbero ad avere un bambino. Senza il finanziamento pubblico ogni coppia dovrebbe spendere dai 5 ai 15 milioni.

CAMPIONE 1 - Lei voterebbe a favore del programma di finanziamento pubblico se le chiedessero un pagamento annuale, una tassa di 5.000 lire?

V771

Print Format: F1
Write Format: F1

Value	Label
0	non so
1	sicuramente no
2	probabilmente no
3	probabilmente si'
4	sicuramente si'

CAMPIONE 2 - Lei voterebbe a favore del programma di finanziamento pubblico se le chiedessero un pagamento annuale, una tassa di 10.000 lire?

V772

Print Format: F1
Write Format: F1

Value Label

0 non so
1 sicuramente no
2 probabilmente no
3 probabilmente si'
4 sicuramente si'

CAMPIONE 3 - Lei voterebbe a favore del programma di finanziamento pubblico se le chiedessero un pagamento annuale, una tassa di 20.000 lire?

V773

Print Format: F1
Write Format: F1

Value Label

0 non so
1 sicuramente no
2 probabilmente no
3 probabilmente si'
4 sicuramente si'

CAMPIONE 4 - Lei voterebbe a favore del programma di finanziamento pubblico se le chiedessero un pagamento annuale, una tassa di 50.000 lire?

V774

Print Format: F1
Write Format: F1

Value Label

0 non so
1 sicuramente no
2 probabilmente no
3 probabilmente si'
4 sicuramente si'

A TUTTI

DOMANDA 8: Ma, quanto sarebbe il contributo annuale che lei personalmente sarebbe disposto a pagare? Per ciascuna delle seguenti alternative, indichi, per favore, se sì o se no, tenendo presente che:

- 2.000 lire

V8.1

Print Format: F1
Write Format: F1

Value Label

1 si'
2 no

- 5.000 lire

V8.2

Print Format: F1
Write Format: F1

Value Label

1 si'
2 no

- 10.000 lire

```
V8.3
          Print Format: F1
          Write Format: F1

          Value    Label

              1    si'
              2    no
```

- 20.000 lire

```
V8.4
          Print Format: F1
          Write Format: F1

          Value    Label

              1    si'
              2    no
```

- 50.000 lire

```
V8.5
          Print Format: F1
          Write Format: F1

          Value    Label

              1    si'
              2    no
```

- 80.000 lire

```
V8.6
          Print Format: F1
          Write Format: F1

          Value    Label

              1    si'
              2    no
```

- 100.000 lire

V8.7
Print Format: F1
Write Format: F1

Value Label

1 si'
2 no

- 200.000 lire

V8.8
Print Format: F1
Write Format: F1

Value Label

1 si'
2 no

DOMANDA 9: Ma, Lei sarebbe favorevole al finanziamento pubblico della FIVET:

Value Label

1 Sì,
2 Sì, ma solo per le fasce a basso reddito
3 Sì, ma con una compartecipazione a carico delle famiglie
4 No

Variabili strutturali

PM6 PANELMEMBER
Print Format: F7
Write Format: F7

SEQUENZA PROGRESSIVO FISSO
Print Format: F2
Write Format: F2

UNIVOCO DENTIFICATIVO DI CASO (pm6 *100000 +sequenza)
Print Format: F16
Write Format: F16

ETACON (età continua)
Print Format: F8.2
Write Format: F8.2

ETA ETA' (per classi)
Print Format: F1
Write Format: F1

Value Label

```
         2    18-29 anni
         3    30-39 anni
         4    40-49 anni
         5    50-59 anni
         6    60 e oltre

ETACL2   ETA' (classi d'età quinquennali)
         Print Format: F3
         Write Format: F3

         Value    Label

             1    18-24 anni
             2    25-29 anni
             3    30-34 anni
             4    35-39 anni
             5    40-44 anni
             6    45-49 anni
             7    50-54 anni
             8    55-59 anni
             9    60-64 anni
            10    >64 anni

SESSO    SESSO
         Print Format: F2
         Write Format: F2
         Missing Values: 0

         Value    Label

             1    maschio
             2    femmina
```

```
ISTR2     SCOLARITA'
          Print Format: F2
          Write Format: F2
          Missing Values: 0

          Value    Label

              1    no tit/lic. elem.
              2    lic. media inf.
              3    dipl. media sup.
              4    laurea


CONPROF3  CONDIZIONE PROFESSIONALE
          Print Format: F2
          Write Format: F2
          Missing Values: 0

          Value    Label

              1    Occupato
              2    In cerca lavoro
              3    Non forze lavoro


CONPROF7  CONDIZIONE PROFESSIONALE
          Print Format: F2
          Write Format: F2

          Value    Label

              1    lav.auton.
              2    lav.dip.
              3    operaio
              4    casa-linga
              5    stu-dente
              6    pensio-nato
              7    in cerca occup.


MCR       MACROAREA
          Print Format: F2
          Write Format: F2

          Value    Label

              1    nord-ovest
              2    nord-est
              3    centro
              4    sud e isole


MCR2      MACROAREA
          Print Format: F6
          Write Format: F6

          Value    Label

              1    Nord ovest
              2    Nord est
              3    Centro
              4    Sud
              5    Isole
```

AMPC AMPIEZZA CENTRO
Print Format: F2
Write Format: F2

Value	Label
1	<=5000
2	5001-20000
3	20001-50000
4	50001-100000
5	>100000

CODREGI REGIONE
Print Format: F3
Write Format: F3

Value	Label
1	Piemonte
2	Valle d'Aosta
3	Lombardia
4	Trentino Alto Adige
5	Veneto
6	Friuli Venezia Giulia
7	Liguria
8	Emilia Romagna
9	Toscana
10	Umbria
11	Marche
12	Lazio
13	Abruzzo
14	Molise
15	Campania
16	Puglia
17	Basilicata
18	Calabria
19	Sicilia
20	Sardegna

CODPROV Codice Istat provincia
Print Format: F2
Write Format: F2

Value	Label
1	Torino
2	Vercelli
3	Novara
4	Cuneo
5	Asti
6	Alessandria
7	Aosta
8	Imperia
9	Savona
10	Genova
11	La Spezia
12	Varese
13	Como
14	Sondrio
15	Milano
16	Bergamo
17	Brescia
18	Pavia
19	Cremona
20	Mantova
21	Bolzano

22 Trento
23 Verona
24 Vicenza
25 Belluno
26 Treviso
27 Venezia
28 Padova
29 Rovigo
30 Udine
31 Gorizia
32 Trieste
33 Piacenza
34 Parma
35 Reggio Emilia
36 Modena
37 Bologna
38 Ferrara
39 Ravenna
40 Forlì
41 Pesaro
42 Ancona
43 Macerata
44 Ascoli Piceno
45 Massa Carrara
46 Lucca
47 Pistoia
48 Firenze
49 Livorno
50 Pisa
51 Arezzo
52 Siena
53 Grosseto
54 Perugia
55 Terni
56 Viterbo
57 Rieti
58 Roma
59 Latina
60 Frosinone
61 Caserta
62 Benevento
63 Napoli
64 Avellino
65 Salerno
66 L'Aquila
67 Teramo
68 Pescara
69 Chieti
70 Campobasso
71 Foggia
72 Bari
73 Taranto
74 Brindisi
75 Lecce
76 Potenza
77 Matera
78 Cosenza
79 Catanzaro
80 Reggio Calabria
81 Trapani
82 Palermo
83 Messina
84 Agrigento
85 Caltanissetta
86 Enna
87 Catania
88 Ragusa
89 Siracusa

```
     90    Sassari
     91    Nuoro
     92    Cagliari
     93    Pordenone
     94    Isernia
     95    Oristano


AMPC2     AMPIEZZA CENTRO
          Print Format: F1
          Write Format: F1

          Value    Label

              1    <=5000
              2    5001-20000
              3    20001-50000
              4    50001-100000
              5    >100000
              6    Capoluogo


CETOVEN   LIVELLO SOCIO ECONOMICO
          Print Format: F6
          Write Format: F6
          Missing Values: 0

          Value    Label

              1    Basso
              2    Medio basso
              3    Medio
              4    Medio alto
              5    Alto


AMFTEO2   AMPIEZZA FAMIGLIA
          Print Format: F2
          Write Format: F2

          Value    Label

              1    1 componente
              2    2 componenti
              3    3 componenti
              4    4 componenti
              5    5 o piu' componenti


CAPOLUO   Comune capoluogo
          Print Format: F1
          Write Format: F1

          Value    Label

              0    No
              1    Si


QUOTIDIA  LETTURA QUOTIDIANI
          Print Format: F1
          Write Format: F1

          Value    Label

              1    tutti i giorni
              2    spesso
              3    saltuariamente
```

```
            4    raramente/mai


RESP      RESP. ACQUISTI
          Print Format: F1
          Write Format: F1

          Value    Label

              0    No
              1    Si'


QUAFAM    Qualifica familiare
          Print Format: F8
          Write Format: F8

          Value    Label

              1    marito/partner
              2    moglie/partner
              3    figlio/figlia
              4    fratello/sorella
              5    zio/zia
              6    cugino/cugina
              7    nipote
              8    nonno/nonna
              9    genero/nuora
             10    cognato/cognata
             11    suocero/suocera
             12    Single
             13    Altro parente
             14    Padre/madre


CAPOF     Ruolo individuo
          Print Format: F1
          Write Format: F1
          Missing Values: 0

          Value    Label

              1    Capofamiglia
              2    Non capofamiglia.
```

```
ETACL4     Eta a 6 classi
           Print Format: F3
           Write Format: F3
           Missing Values: 99

           Value    Label

               1    18-24 anni
               2    25-34 anni
               3    35-44 anni
               4    45-54 anni
               5    55-64 anni
               6    >64 anni


ELE96PTO   VOTO PROPORZIONALE (ricordo+rilevato)
           Print Format: F3
           Write Format: F3

           Value    Label

               1    FI
               2    AN
               3    CCD,CDU
               4    Pannell
               5    PDS
               6    RC
               7    Verdi, Rete
               8    Dini, Segni, SI, AD
               9    Popol-U.dem.+Labur.
              10    Lega Nord
              11    Altri
              12    Non Voto


ELE96M     VOTO MAGGIORITARIO (rilevato50)
           Print Format: F2
           Write Format: F2

           Value    Label

               1    CENTRO-DESTRA
               2    CENTRO-SINISTRA
               3    Lega Nord
               4    Pannell
               5    Altri
               6    Non voto


MESSA      Esclusi i matrimoni e i funerali, con quale frequenza partec
           Print Format: F1
           Write Format: F1

           Value    Label

               1    Mai
               2    1-2 volte l'anno
               3    Piu' volte l'anno
               4    1 volta al mese circa
               5    2-3 volte al mese
               6    Ogni settimana
               7    Piu' volte la settimana
```

```
RMESSA    FREQUENZA ALLA MESSA
          Print Format: F8
          Write Format: F8

          Value    Label

              1    mai
              2    1\2 volte anno
              3    piu' volte anno
              4    1\3 volte mese
              5    ogni sett.

RV6       SCELTA ATTUALE DI VOTO (con non voto)
          Print Format: F8.2
          Write Format: F8.2

          Value    Label

           3,00    Forza Italia
           4,00    All.Naz
           5,00    CCD
           6,00    Lista-Pannel.
           7,00    UDR
           8,00    Dem. di Sinis.
           9,00    Rif.Com
          10,00    Com. d'Italia
          11,00    Verdi
          12,00    Rinnov.Ital.
          13,00    Ulivo
          14,00    L'Ital.dei Val
          15,00    PPI
          16,00    Lega - Nord
          17,00    MS Fiam. Tricol
          18,00    Social-Italian
          19,00    Patto  Segni
          20,00    Altri
          21,00    Bianca/nulla/N.V.

RV9       VOTO MAGGIORITARIO ATTUALE (con non voto)
          Print Format: F8.2
          Write Format: F8.2

          Value    Label

           1,00    Polo-Liberta'
           2,00    L'Ulivo
           3,00    Rif.Com.
           4,00    Lega-Nord
           5,00    Lista  Pannel.
           6,00    Altro
           7,00    Bianca/nulla/N.V.
```

```
V5R2      Autocollocazione (con non so)
          Print Format: F8.2
          Write Format: F8.2

          Value     Label

           1,00     sini-stra
           2,00     centro-sin.
           3,00     centro
           4,00     centro-destra
           5,00     destra
           6,00     non so


PA13      INTERESSE ALLA POLITICA
          Print Format: F8.2
          Write Format: F8.2

          Value     Label

            ,00     non so
           1,00     per nulla
           2,00     poco
           3,00     abbastanza
           4,00     molto


DATANAS   Datanasci aa.mm.gg.
          Print Format: EDATE8
          Write Format: EDATE8


PESO4     peso senza politiche
          Print Format: F13.10
          Write Format: F13.10


PESO4C1   peso senza politiche (campione 1)
          Print Format: F13.10
          Write Format: F13.10


PESO4C2   peso senza politiche (campione 2)
          Print Format: F13.10
          Write Format: F13.10


PESO4C3   peso senza politiche (campione 3)
          Print Format: F13.10
          Write Format: F13.10


PESO4C4   peso senza politiche (campione 4)
          Print Format: F13.10
          Write Format: F13.10


NFIGLI    Numero di figli
          Print Format: F4
          Write Format: F4
```

```
V5      In politica di solito si parla di 'SINISTRA', 'CENTRO' e 'DE
        Print Format: F2
        Write Format: F2

V6      Se oggi ci fossero delle nuove elezioni, quale partito voterà
        Print Format: F2
        Write Format: F2

        Value    Label

            3    Forza Italia (Berlusconi)
            4    Alleanza Nazionale (Fini)
            5    CCD (Casini)
            6    Lista Pannella (Pannella)
            7    UDR (Cossiga-Mastella)
            8    Democratici di Sinistra (ex PDS-Veltroni)
            9    Rifondazione Comunista (Bertinotti)
           10    Comunisti d'Italia (Cossutta)
           11    Verdi (Manconi)
           12    Rinnovamento Italiano (Dini)
           13    l'Ulivo (Prodi)
           14    Italia dei valori (Di Pietro)
           15    PPI (Marini)
           16    Lega Nord (Bossi)
           17    Movimento Sociale - Fiamma Tricolore (Rauti)
           18    Socialisti Italiani di Boselli
           19    Patto Segni (Segni)
           20    altri
           21    scheda bianca
           22    scheda nulla
           23    non andrei a votare

V9      Se oggi ci fossero delle nuove elezioni, lei quale coalizion
        Print Format: F1
        Write Format: F1

        Value    Label

            1    Polo per le Liberta'
            2    L'Ulivo
            3    Rifondazione Comunista
            4    Lega Nord
            5    Lista Pannella
            6    altro
            7    scheda bianca
            8    scheda nulla
            9    non andrei a votare
```

Questionnaire
(English translation)

QUESTION 1: Are you acquainted with any couples who desire to have babies and cannot have them?

Yes
No

QUESTION 2: Did you know In-Vitro-Fertilisation before?

No, It is the first time I hear of it
Yes, I have heard of it from the Tv and/or newspapers
Yes, I have heard of it from relatives/friends
Yes, I have got informed through specialised readings
Yes, I have got informed from the doctor

QUESTION 3: How would you rate your knowledge about In-Vitro-Fertilisation?
Very limited
Relatively limited
Sufficient
Relatively wide
Very wide

QUESTION 4: Have you ever met someone who used IVF or other ARTs?
Yes
No

QUESTION 5: Imagine that you have been married for a few years and you have been unsuccessfully trying to have a baby. Imagine you are told to use IVF and that it has a 30% chance to be successful. Would you personally try IVF?
Do not know
No, definitely not
No, probably not
Yes, probably
Yes, definitely

QUESTION 6: Would you personally try IVF if it were free?
Yes
No

QUESTION 6A: Would you personally try IVF if you had to pay 1 Million Italian Liras?
Yes
No

QUESTION 6B: Would you personally try IVF if you had to pay 5 Million Italian Liras?
Yes
No

QUESTION 6C: Would you personally try IVF if you had to pay 10 Million Italian Liras?
Yes
No

QUESTION 6D: Would you personally try IVF if you had to pay 20 Million Italian Liras?
Yes
No

QUESTION 6E: Would you personally try IVF if you had to pay 50 Million Italian Liras?
Yes
No

Sample splitted into 4 sub-samples

QUESTION 7: Imagine that there is a referendum in order to decide whether to fund IVF with public money. Each year public funding would be available to about 30,000 couples who are advised to use IVF. Out of these couples about 10,000 would have a baby. Without public funding, couples requiring IVF should spend between It. Liras 5 and It. Liras 15 million for the procedure.

Sub-sample 1 – Would you vote in favour of funding IVF with public money if you were asked an annual payment, that is a tax, of Italian Liras 5,000?

Do not know
No, definitely not
No, probably not
Yes, probably
Yes, definitely

Sub-sample 2 – Would you vote in favour of funding IVF with public money if you were asked an annual payment, that is a tax, of Italian Liras 10,000?

Do not know
No, definitely not
No, probably not
Yes, probably
Yes, definitely

Sub-sample 3 – Would you vote in favour of funding IVF with public money if you were asked an annual payment, that is a tax, of Italian Liras 20,000?

Do not know
No, definitely not
No, probably not
Yes, probably
Yes, definitely

Sub-sample 4 – Would you vote in favour of funding IVF with public money if you were asked an annual payment, that is a tax, of Italian Liras 50,000?

Do not know
No, definitely not
No, probably not
Yes, probably
Yes, definitely

All

QUESTION 8: And, would you try IVF if.......Please, for each alternative indicate yes or not
- 2.000 liras
- 5,000 liras
- 10,000 liras
- 20,000 liras
- 50,000 liras
- 80,000 liras
- 100,000 liras
- 200,000 liras

QUESTION 9: But, would you in favour to a public program providing IVF to infertile couples?

Yes
Yes, but only for low incombe couplet
Yes, but with a co-payment
No