Using MCDA (Multi-Criteria Decision Analysis) to prioritise publicly-funded health care

Trudy Sullivan

A thesis submitted for the degree of

Doctor of Philosophy

at the University of Otago, Dunedin,

New Zealand.

October 2012

~ Abstract ~

New Zealand, like many other countries, is grappling with the problem of how to allocate limited resources across a range of health and disability support services at a time when demand for health care continues to grow faster than health budgets. It is becoming increasingly important for decision-makers to adopt robust processes for setting priorities so that limited health resources are allocated efficiently, effectively and transparently. In my thesis I use multi-criteria decision analysis (MCDA) to build a framework (at the meso-level of health care funding) which can be used by decision-makers to assist them in priority-setting.

Potential criteria, elicited from six focus groups (including members of the public, private and public health care providers, health professionals and policy makers), are combined with advice from health experts and criteria from comparable studies in the literature to establish six prioritisation criteria: 'need', 'individual benefit', 'societal benefit, 'age', 'lifestyle' and 'no alternative treatment'. An online decision survey implemented through 1000Minds software (Ombler & Hansen 2012) and the PAPRIKA[1] method (Hansen & Ombler 2008) is used to determine the relative importance of the criteria. According to the results of a 'test re-test', the decision survey accurately captures the preferences of respondents.

The results of the decision survey reveal that 'need' and 'individual benefit' are the most important prioritisation criteria, and though patients are unlikely to be prioritised according to their age or lifestyle (because of discrimination), greater preference is shown for 'age' and 'lifestyle' compared to 'societal benefit' and 'no alternative treatment'.

Regression analysis (including the application of a fractional multinomial logit model) and cluster analysis are used to determine whether the demographic characteristics of respondents can predict preferences. Several relationships are found. For example, health care workers, respondents on low incomes and Maori place more importance on 'need' (relative to the other criteria) compared to respondents who do not work in health care, respondents on middle or high incomes and non-Maori. Though several statistically significant results are found, it appears that overall the variation in preferences is largely due to the idiosyncrasies of respondents and not to particular demographic characteristics.

[1] Potentially All Pairwise RanKings of all possible Alternatives.

The criteria weights from the random sample are then brought together with cost and other additional factors in a prioritisation framework. With the aid of a Value for Money (VfM) chart and associated budget allocation table, decision-makers can consider all the prioritisation variables in a transparent and consistent way. The framework can be used as a communication tool, to allocate fixed budgets across a range of services, to keep track of previous decisions or to re-allocate resources when the budget has been cut.

The framework developed in this thesis illustrates how health care can be prioritised at the meso-level of health care funding in New Zealand. Ultimately it is up to the decision-makers to choose which treatments to fund, but if decisions are made explicitly within a transparent and robust framework that includes all relevant considerations (including the preferences of key stakeholders), then there is likely to be more acceptance in the outcome.

I presented a seminar on "Priority-setting in a publicly-funded health system" at:

i) PHARMAC, Wellington, 18 May 2012
ii) The Department of Pharmacology and Toxicology, University of Otago, 7 June 2012
iii) The Department of Preventive and Social Medicine, University of Otago, 15 August 2012
iv) The Department of Bioethics, University of Otago, 3 September 2012

I will present a workshop entitled "Applying MCDA (Multi-Criteria Decision Analysis) to Health Care Prioritisation" at The New Zealand Hospital Tertiary Services National Conference on 22 November 2012.

~ Acknowledgements ~

Thank you to my supervisors (Associate Professor Paul Hansen, Dr Paul Thorsnes and Professor Rob Lawson), and to my family, friends and colleagues, who inspired, encouraged, motivated, helped and supported me in writing this thesis.

I would also like to thank the following people for their advice, guidance or assistance:

Vignette development: Dr Melissa Audeau, Mr Richard Bunton, Associate Professor Brian Cox, Dr Nigel Dickson, Associate Professor Wayne Gillett, Professor John Highton, Dr Annette Neylon, Dr Patricia Priest, Associate Professor Jean-Claude Theis, Professor Rob Walker (Dunedin School of Medicine), Barb Long (Royal NZ Plunket Society Inc), Dr Jill McIlraith (Sexual Health Clinic, Dunedin Hospital), Dr Ian Sheerin (Christchurch School of Medicine), Dr Andrew Harrison, Associate Professor Stewart Mann (Wellington School of Medicine).

Rachel Grocott, Rico Schoeler and Matthew Poynton (PHARMAC); Marianne Hannagan (Lysosomal Diseases New Zealand); Dr Ofra Golan (The Gertner Institute for Epidemiology & Health Policy Research, Israel); Professor Peter Herbison, Dr Murat Genc, Professor Jeff Smith, Aurelia Lepine and Alexis Le Nestour (University of Otago), Professor Toni Ashton (University of Auckland).

Participants in the six focus groups and respondents in the pre-tests, pilot study, snowball sample, random sample and health services researchers' sample.

I am grateful to everyone for their assistance. However, the work is my own and I am responsible for the content of this thesis.

~ Table of Contents ~

Chapter 3: Focus groups

Chapter 4: Establishing Criteria for the Decision Survey

Chapter 5: The decision survey and sample groups

Chapter 6: Respondents' preferences for the six criteria

Chapter 7: Do respondents' characteristics predict preferences?

Chapter 8: Cluster analysis

~ List of tables ~

~ List of figures ~

~ Chapter 1 ~

Introduction

1.1 Introduction

New Zealand, like most other countries, is grappling with the problem of how to allocate limited resources across a range of health and disability support services at a time when demand for health care continues to grow faster than health budgets. An aging population[2] only adds to the escalating health care costs. The media, lobby groups, pharmaceutical companies and increased access to the internet have all contributed to a growing awareness of new treatments than can offer better outcomes (Menon & Stafinski 2007). It is becoming increasingly important for decision-makers to adopt robust processes for setting priorities so that limited health resources are allocated efficiently, effectively and transparently.

In this thesis I use a multi-criteria decision analysis (MCDA)[3] to develop a prioritisation framework, applicable at the meso-level of health care funding, which incorporates the preferences of a cross-section of the New Zealand public. By including key stakeholders in the process – members of the general public, health experts, private and public health care workers, health services researchers –relevant considerations including the preferences of the general public can be incorporated into a framework which will ultimately be used as a tool to assist decision-makers in priority-setting.

The chapter begins with a brief background on health care prioritisation in New Zealand's public health sector followed by an explanation of the MCDA approach used in this thesis. The chapter concludes with an overview of the thesis and a glossary of some of the key terms.

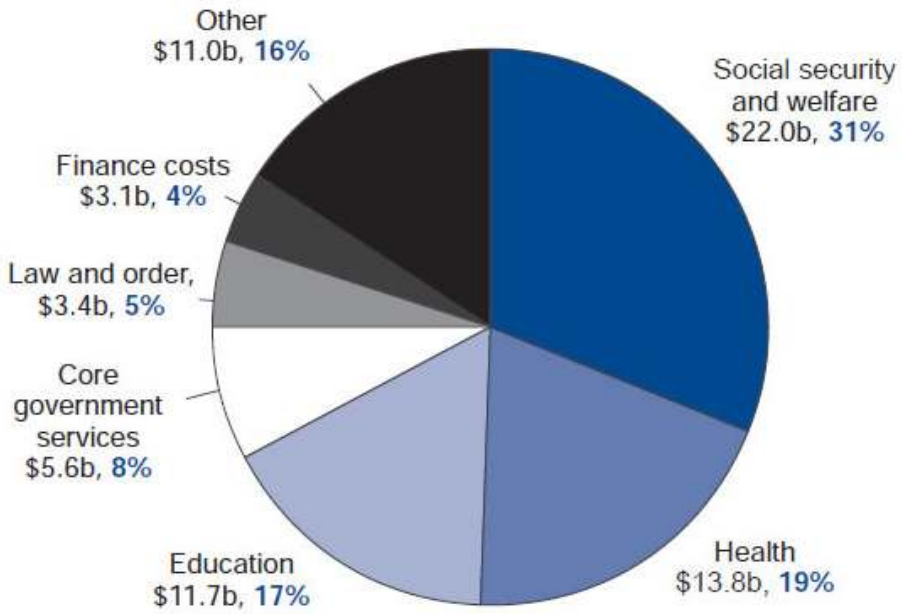
1.2 Health care prioritisation in New Zealand's public health sector

Prioritisation[4] within the public health sector can be thought of as occurring at three levels: macro, meso and micro (Logan et al. 2004). The macro-level refers to the overall budget decisions made by the Government. For example, as illustrated in Figure 1.1, in the 2010/11

[2] The 65+ age group is projected to make up over 25% of New Zealand's population from the late 2030s, compared with 12% in 2005 (Statistics New Zealand).
[3] MCDA is explained in Section 1.2.
[4] The terms prioritisation' and 'priority-setting' are used interchangeably throughout this thesis.

financial year $13.8 billion was spent on health care in New Zealand, which equates to 19% of the total spending on core Crown spending and approximately 7% of GDP (Statistics New Zealand, 2011).

Figure 1.1: New Zealand core Crown spending for 2010/2011 year



Source: The New Zealand Treasury (2011)

Allocating the health budget across a range of health and disability services occurs at the meso-level. Health agencies within New Zealand, including the Ministry of Health, 20 District Health Boards, the Pharmaceutical Management Agency (PHARMAC)[5] and independent providers of health and disability services determine what health and disability services are to be funded from the overall budget. In 2007/2008 a set of health targets, designed to improve the health sector's performance and agreed to by the Ministry of Health and the District Health Boards, was introduced. The health targets are reviewed annually to reflect priority health areas determined by the Government. The health targets for 2011/2012 are: shorter stays in emergency departments, improved access to elective surgery, shorter waits for cancer treatment, increased immunisation, better help for smokers to quit and better diabetes and cardiovascular services.

[5] PHARMAC was established in 1993 as part of the New Zealand medicines system "to find new and effective ways to manage expenditure growth, while also obtaining the best health outcomes for the New Zealand population" (Grocott 2009, p 181).

The micro-level of priority-setting refers to the clinical decision-making made by health professionals; that is, determining which patients should receive treatment and what that treatment entails. The three levels of priority-setting are part of a continuum of decision-making (Logan et al. 2004). For example, the priorities set by a funding authority will impact on the decisions made by a doctor in prioritising his or her patients.

In this thesis, the main emphasis is on prioritising health care at the meso-level. Deciding which health treatments and/or programmes to fund is a complex issue. An increasingly popular approach used by decision-makers to allocate resources is MCDA (Devlin & Sussex 2010). A MCDA approach is used in this thesis to develop a prioritisation framework and is explained in the next section.

1.3 Multi-Criteria Decision Analysis (MCDA)

Deciding who receives priority when allocating scarce health resources is difficult and challenging. Should the sickest members of our society be treated first? Should young people receive priority over old people? Should we spend more on palliative care? MCDA or a MCDA-based approach confronts these inevitable trade-offs. MCDA helps decision-makers to consider multiple conflicting factors or 'criteria' in a rational and consistent way (Baltussen 2006).

MCDA is based on the premise that any good or service (in this case, health treatments and/or programmes) can be described by its characteristics (criteria), and the extent to which an individual values the health treatments and/or programmes depends on the individual's preferences for those characteristics (Ryan 2004).

As well as being linked to utility theory, many aspects of MCDA are related to voting oriented social choice theory and welfare economics (Scott & Antonsson 2000). Social choice theory seeks to defines ways in which the preferences of individuals can be aggregated in order to reach a collective decision (or in other words, to construct a social welfare function). Similarly with MCDA, the preferences of individuals for a defined set of criteria are elicited and aggregated, in order to produce a preference ranking of possible alternatives (Scott & Antonsson 2000, Figueira et al. 2005).

The problem common to both social choice theory and MCDA is that there is more than one decision-maker. The rankings of alternatives by individuals need to be combined in a way that incorporates fairly, the preferences of the individuals involved in the decision. To illustrate the problem, consider the following. Three voters rank three alternatives, A, B and C, as follows: A≿B≿C,[6] B≿C≿A and C≿A≿B. When considered pairwise (and using transitivity), a paradox arises because as a group, A is preferred to B, B is preferred to C and C is preferred to A. The resulting social order provides no basis on which to make a decision (Scott & Antonsson 2000). It is this problem that is explored within social choice theory.

In his well-known work on social choice theory, Arrow (1963) used a set of axioms relating to voting procedures including Pareto inclusiveness and independence of irrelevant alternatives (IIA)[7] to investigate whether individual ordinal preferences can reasonably be aggregated into social preferences. Arrow found that there is no consistent, equitable method for aggregating individual preferences – that dictatorship is the only social choice rule capable of socially ordering individual preferences (Scott & Antonsson 2000, Hansen 2002).[8]

However, the range of social choice rules can be expanded if one (or more) of the axioms is relaxed, for example, the IIA assumption, or if additional information relating to individuals' preferences (i.e. more than just their ordinal rankings) is included in the ranking process (Hansen 2002). With social choice theory, each option is worth the same (i.e. preferences are ordinally measurable) and there is no interpersonal comparability (i.e. the preferences of one individual cannot be compared with another). Similar to social choice theory, MCDA methods also require decision-makers to express ordinal preferences. The 'paradox' explained above could also occur with MCDA. For example, consider three alternatives, A, B and C each of which has three criteria, a, b, and c. If a, b and c have equal values it is possible that A≿B≿C, B≿C≿A and C≿A≿B with respect to a. However, additional information on a, b, and c (such as strength of preference) will enable a social ranking of A, B and C. With MCDA each option (or criterion) can be weighted using various scoring methodologies (discussed in Section 2.7). Cardinal scores are created from ordinal rankings thereby facilitating the

[6] Where A≿B means that A is at least as preferable as B.
[7] The five axioms are collective rationality, universal domain, Pareto inclusiveness, independence of irrelevant alternatives and non-dictatorship. For a full explanation of Arrow's Impossibility Theorem, see any textbook relating to social choice theory or Arrow (1963).
[8] This well-known objection to combining ordinal preferences into a single order refers to Arrow's General Possibility Theorem, commonly known as Arrow's Impossibility Theorem or Arrow's Theorem.

aggregation of individual preferences and avoiding the individual orderings problems associated with Arrow's Impossibility Theorem (Dolan 1998).[9]

The potential trade-off between efficiency and equity is often modelled within the traditional framework of welfare economics. In MCDA, the relative importance of competing criteria indicates amongst other things, the degree to which society wishes to maximise total population health (efficiency) and/or to look after the sickest members of society (equity).[10]

A typical MCDA process, and the process followed in this thesis, involves identifying the health treatments and/or programmes that need to be prioritised, deciding who should be involved in the MCDA process, establishing the criteria by which the health treatments will be prioritised and determining the relative importance of the criteria. Criteria weights (representing the relative importance of each of the criteria) can then be used by decision-makers to assist them in prioritising services. For example, criteria weights can be summed across all the criteria for each health treatment to obtain an overall score. The overall scores can then be used to rank the health treatments and/or programmes. A MCDA process ensures consistency and transparency. In addition, when key stakeholders are involved throughout the process, there is likely to be greater acceptance of the outcome.

An outline of my thesis is presented in the next section.

1.4 Thesis outline

There are two parts to this thesis. The first part of the thesis uses qualitative methods to determine the relevant criteria by which health treatments and/or programmes can be prioritised. The second part of the thesis uses quantitative methods to determine and analyse the relative importance of the criteria. A brief summary of each of the Chapters is given in this section.

Chapter 2: The thesis begins with an overview of the different approaches countries have used for prioritising health care. Some of the methods that are commonly used to develop prioritisation frameworks, including MCDA are described. The steps of an MCDA process

[9] For example, the PAPRIKA method, discussed in Section 2.7 produces 'point values' or 'weights', (representing the relative utilities of each level of each criterion), by solving the inequalities (strict preference) and equalities (indifference) resulting from a set of ordinal preferences, with a linear programme.

[10] The competing objectives of efficiency and equity reflect utilitarianism (maximising utility) and Rawlsianism (minimising inequality) respectively.

are explained with particular emphasis on the input of key stakeholders in the process and the different scoring methodologies available for calculating criteria weights.

Chapter 3: An important part of the MCDA process is to determine the appropriate criteria for prioritising health services. In this thesis, focus groups, with the aid of health 'vignettes', are used to discover the criteria by which health services *should* be prioritised. The various qualitative methods that can be used to elicit criteria are discussed in this chapter before focussing on the establishment of the focus groups. The chapter concludes with a summary of the results of the six focus group meetings.

Chapter 4: This chapter starts with a review of a number of studies exploring health care prioritisation. The suggested criteria from the focus group meetings are then considered together with potential criteria from the literature and other relevant information to construct appropriate criteria and levels for a decision survey (explained below).

In the second part of the thesis an online choice-based survey implemented through 1000Minds software[11] (Ombler & Hansen 2012) is used to elicit the preferences of the general public with regard to the relative importance of each of the criteria. In the survey, respondents are asked to trade-off one characteristic of health care with another, by choosing which of two imaginary patients to treat first (assuming all other characteristics are equal). The PAPRIKA (Potentially All Pairwise RanKings of all possible Alternatives) scoring method[12] (Hansen & Ombler 2008) is used to obtain the respondents' relative preferences for the criteria (i.e. the criteria weights). One of the main advantages of using the PAPRIKA method is that individual criteria weights can be calculated for *every* individual.[13] The willingness to trade-off aspects of health care can then be compared between individuals and between groups of individuals.

Chapter 5: In this chapter the process of developing the decision survey and establishing the sample groups is described. Several groups of respondents completed the decision survey including a pre-test group, a pilot sample, a 'snowball' sample and a random sample. Several issues relating to sample selection and the response rate are also discussed.

[11] www.1000minds.com
[12] A full explanation of the PAPRIKA method is given in Chapter 2.
[13] Typically with other scoring methods, the criteria weights represent preferences aggregated over the entire sample. In Chapter 2 several different scoring methods are explained and compared to the PAPRIKA method.

Chapter 6: The criteria weights for the sample groups are presented and discussed. A summary of the comments made by random sample respondents is included to provide some insight as to how they made their decisions. The chapter concludes with an evaluation of the validity and reliability of the survey including the results of a 'test re-test'.

Chapters 7: The data from the random sample are analysed using regression analysis. The results of the analyses are presented in this chapter. Several relationships are found between the demographic characteristics of the respondents and their relative preferences for individual criteria.

Chapter 8: Respondents from the random sample are clustered into groups based on the similarity of their criteria weights and the resulting cluster groups examined to determine whether respondents who have similar criteria weights share common demographic characteristics. Several relationships are found and these results are presented. In addition, the 14 health vignettes used in the focus groups are used in combination with the criteria weights from each cluster and some 'arbitrarily assigned' weights to illustrate how treatment rankings differ depending on whose criteria weights are used.

Chapter 9: In this chapter the criteria weights from the random sample are brought together with cost and other 'additional factors'. A value for money (VfM) chart and associated budget allocation table are used to illustrate how decision-makers can consider all the prioritisation variables in a transparent and consistent way when allocating funding across a range of health treatments and/or programmes.

Chapter 10: The main findings of the thesis are summarised in Chapter 10. The benefits and limitations of the proposed prioritisation framework are also discussed together with recommendations for future research.

1.5 Glossary

Below is a glossary of some of the key terms used throughout the thesis.

Criterion/criteria

A criterion is a characteristic, factor or attribute on which a decision can be based.

Decision survey

An online choice-based survey (implemented through 1000Minds software (Ombler & Hansen 2012) and the PAPRIKA scoring method (Hansen & Ombler 2008)) is used to estimate the preferences of the general public with respect to the six criteria. Respondents are presented with a series of hypothetical choices, each of which involves two imaginary patients who differ in only two characteristics. Each choice requires a respondent to trade-off one criterion for another by choosing which patient to treat first.

Health-related quality of life (HRQoL)

HRQoL is a patient-reported outcome that relates to the quality of life of a patient in regard to health or health care. It encompasses physical, mental and emotional factors; for example, ability to perform daily tasks, to work or to be free of depression.

Health technology assessment (HTA)

New health technologies are evaluated before they become part of clinical practice. This is known as 'health technology assessment'. Traditionally HTA has focussed on the efficacy, effectiveness and cost-effectiveness of a new technology; however, other considerations including social, legal, ethical and political factors are now commonly included in HTA.

Incremental cost effectiveness ratio (ICER)

An ICER is used to compare the cost per Quality Adjusted Life Year (QALY) (explained below) of two competing health interventions. The ratio is derived by dividing the additional cost of a new treatment by the additional QALYs of the new treatment compared with a comparable alternative treatment.

Multi-criteria decision analysis (MCDA):

MCDA approaches are used to assist individuals and/or groups in making complex decisions, involving multiple conflicting considerations ('criteria'), in an explicit and transparent way. A MCDA approach identifies the relevant criteria to be considered and determines the influence these multiple criteria have in the decision-making process.

Potentially All Pairwise RanKings of all possible Alternatives (PAPRIKA)

In this thesis, the PAPRIKA scoring method, implemented through 1000Minds software, is used to derive estimates of utility associated with the criteria (that is, the relative importance of the criteria).

Quality-Adjusted Life Year (QALY)

A QALY combines length of life and quality of life into one measure. The number of QALYs gained from an intervention is calculated by multiplying the duration of survival by a utility weight representing a patient's health-related quality of life (HRQoL). For example, one year in perfect health is equal to 1.0 QALY, whereas one year in a health state with a HRQoL valued at 0.5 is equal to 0.5 QALY.

~ Chapter 2 ~

Setting the scene

2.1 Introduction

Deciding how to allocate limited resources across a range of competing health services is a complex and difficult process. As most health care in New Zealand and in many other countries is publicly-funded, there is a growing need for prioritisation decisions to be made in an explicit, transparent and consistent way.

The chapter begins with an overview of the different prioritisation approaches that have been used internationally. Various methods used to develop prioritisation frameworks are then discussed, including multi-criteria decision analysis (MCDA), the method used in this thesis. The various steps of MCDA are explained with particular emphasis on whose views should be considered when establishing a prioritisation framework and the different 'scoring' methods that can be used.

The chapter concludes with an outline of the PAPRIKA scoring method together with an explanation of *why* the PAPRIKA method is used in this thesis.

2.2 Health care prioritisation – international experience

Priority-setting of health care can be done implicitly or explicitly. Traditionally an *implicit* approach to health care prioritisation in publicly-funded health systems has been adopted whereby decisions are largely based on historical resource allocations and the rationale for these decisions is not made clear (Logan et al. 2004,Teng et al. 2007). Since the late 1980s, however, an increase in demand for health care coupled with limited public health budgets, has increased the desirability of *explicit* prioritisation whereby specific processes and criteria designed for priority-setting are made publicly available (Logan et al. 2004, Baltussen et al. 2007, Persad et al. 2009).

Establishing explicit priority-setting processes has been approached in a variety of ways. For instance, Norway, the Netherlands and Sweden have mostly defined the principles that ought to be included in a prioritisation process. On the other hand, New Zealand, Israel, the United

Kingdom (UK) and the United States (US) state of Oregon have been more explicit in their priority-setting approaches by defining the health services that will be publicly-funded and/or establishing clinical guidelines (Sabik & Lie 2008).

The priority-setting approaches of these countries and the US state of Oregon are briefly outlined in this section.

2.2.1 Norway

Norway was the first country to attempt explicit priority-setting of health care (Calltorp 1999). In 1987 the Lønning Commission, set up to explore health care prioritisation at the national and local levels, proposed a priority system based on a single principle – the severity of a condition – with five groups ranging from lifesaving treatments to treatments that have no documented effects. This system was to be used as a guide for funding various treatments (Sabik & Lie 2008). However, the framework was difficult to implement and in 1997 a second commission was appointed, the Lønning Commission II, to review the methodology for establishing priorities (Kenny & Joffres 2008). The Lønning Commission II recommended that additional principles – the expected benefit from treatment and cost-effectiveness – should also be considered alongside severity of the condition when categorising treatments. Four priority groups were introduced to categorise health treatments: 'basic services', 'additional services', 'low priority services' and 'no priority services' (Calltorp 1999). As a result, clinicians set priorities within their specialties and, in turn, this information is used by decision-makers in funding decisions. The primary focus of the Lønning II Model is transparent decision-making with a clear link between clinical and political priority-setting (Logan 2004, Sabik & Lie 2008). In 2007 the Norwegian Council for Priority Setting and Quality Improvement was established to provide advice on priority-setting, improve interaction between local and national levels and to encourage transparency in the prioritisation process.

2.2.2 The Netherlands

Health care prioritisation became a prominent issue in the Netherlands in the 1980s. The main goal was to define a publicly-funded basic insurance package in an effort to reform health care expenditure. A controversial proposal by the Dekker Committee suggested that a

universal care package be limited to 85% of the services provided at the time. The proposal was not implemented because of public opposition (Kenny & Joffres 2008).

In 1991 the Committee on Choices in Health Care produced the 'Dunning report'. The Committee suggested that only necessary services be publicly provided and that non-essential services be cut from the package. Four hierarchical priority principles (criteria) were proposed to achieve this objective: 'necessity' (care which is necessary to maintain or restore health), 'effectiveness' (treatment has to be proven and documented), 'efficiency' (efficient delivery based on the results of cost effectiveness studies) and 'individual responsibility' (when an individual cannot afford to pay for treatment) (Stolk & Poley 2005). At first the proposed criteria received broad public support. However, there have been problems in applying the criteria in health care decision-making due to inconsistent interpretation of the criteria and the 'pass/fail' nature of the process (Stolk & Poley 2005).[14]

2.2.3 Sweden

The Swedish Parliamentary Priorities Commission was appointed in 1992 to explore priority-setting of health care. The Commission explicitly rejected defining health services that should or should not be funded and instead outlined three hierarchical principles for priority-setting: human dignity (everyone is equally valuable), need and solidarity (people with the greatest need should be treated first) and cost-efficiency (only to be used when considering treatments for the *same* condition) (Logan et al. 2004, Sabik & Lie 2008). The Commission also defined five priority groups, based on the type of disease or treatment, to be used as a general guide by decision-makers at the clinical, management and political levels (Calltorp 1999). Public feedback was sought on the Commission's recommendations in 1994 and the Swedish Parliament ratified the Commission's proposals in 1997 (Carlsson 2010). A follow-up report in 2000 concluded that the proposals had some impact on priority-setting at the national and local levels but that there was wide variation in priority-setting between providers and that further work was needed. In 2001 the Swedish Government and the Federation of Swedish County Councils collaborated to form a national knowledge centre for priority-setting of health care and social services known as The National Center on Priority Setting in Health Care. The Center provides a consultancy and educational service with a focus on developing transparent prioritisation processes in health care (Carlsson 2010).

[14] As the criteria are hierarchical, a health treatment needs to 'pass' the first criterion of 'necessity' before it will be judged on the second criterion of 'effectiveness' and so on. If a treatment 'fails' any criterion, it will not be publicly-funded.

2.2.4 US state of Oregon

Although Norway was the first country to attempt to explicitly set priorities in health care, the US state of Oregon was the first to attempt to formulate a list of prioritised health services (Logan et al. 2004). At the end of the 1980s, as a result of rising medical costs and a growing number of people who were unable to afford health care because they were uninsured or they did not qualify for federal assistance (Medicaid), Oregon attempted to develop a transparent process for prioritising state-funded medical services. The goal was to increase coverage of Medicaid from 58% of Oregonians below the federal poverty line (FDL) to *all* Oregonians below the federal poverty line by limiting coverage to a basic bundle of health care services (Sabik & Lie 2008).

A working party, established to explore what health services Oregon's Medicaid programme should cover, developed some guiding principles. These included: access to a basic level of care must be universal, society is responsible for financing care for poor people, and a basic level of care must be defined through a public process (Crawshaw et al. 1990). In 1989 the Health Services Commission[15] was charged with developing a prioritised list of services that would be covered by Medicaid. The public was consulted by way of public hearings, town meetings and a telephone survey. The process of establishing the prioritised list was contentious with the first list being rejected by the public. However, the list of prioritised services was modified and has continued to be modified over the years as circumstances have changed, such as cuts in funding. Though the priority-setting process in Oregon has been challenging it has been successful in increasing the number of people with access to Medicaid services (Ham 1997).

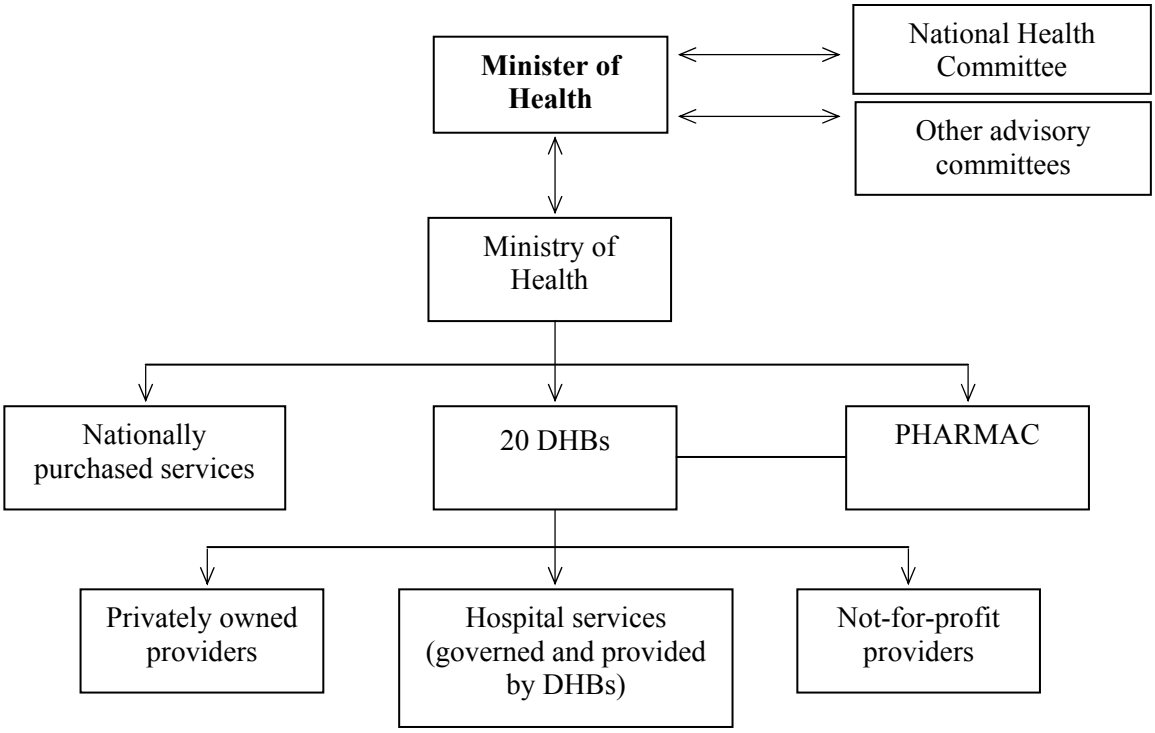
2.2.5 New Zealand

New Zealand's health sector has undergone several major reforms in the last two decades. In its first major restructure in 1993, 14 Area Health Boards were abolished and four regional health authorities (RHAs) were established to purchase all health and disability services for citizens in their region. The main objectives of splitting the roles of purchaser and provider were to increase efficiency, and to improve access to health care. However, the intended outcomes did not eventuate leading to further restructuring (Ashton 1999). In the mid 1990s

[15] The Health Services Commission consisted of a panel of experts made up of five doctors, four consumer members, a public health nurse, and a social worker. As part of their brief they asked 560 people to prioritise a list of 16 health states.

the four RHAs were replaced by one national Health Funding Authority (HFA) which was responsible for purchasing the full range of health and disability services for New Zealanders. The HFA was dissolved in 2001. The Ministry of Health (MOH) took over funding responsibilities and 21 District Health Boards (DHBs) were established to purchase and provide services for their geographical populations (Bloomfield 2003). The current structure of the health and disability sector in New Zealand is illustrated in Figure 2.1.

Figure 2.1: Structure of the health and disability sector in New Zealand



Around the time of the first restructure a Core Services Committee[16] was appointed by the Government to increase public awareness about health care prioritisation and to determine an acceptable way of defining which core health and disability support services should be publicly-funded (Kenny & Joffres 2008). However, after wide public consultation, the Core Services Committee came to the realisation that defining a list of publicly-funded health services would be too difficult and contentious (Gauld & Derrett 2000). Instead, the committee, re-named the National Health Committee (NHC), developed four guiding principles for prioritising health services: effectiveness, equity, acceptability and efficiency (Ashton et al. 2000, Logan et al. 2004).

[16] A full account of the establishment of the Core Services Committee and its objectives is in Cummings (1994).

In 2011 the NHC was further re-configured in order to focus on improving the prioritisation of new and existing health technologies, including systems and models of care (NHC 2012). Currently the NHC uses 11 criteria to guide their decision-making in terms of investment and disinvestment: clinical safety and effectiveness, health and independence gain, materiality, feasibility, policy congruence, equity, acceptability, cost effectiveness (value for money), affordability, risk, and other criteria as the NHC thinks fit.

In order to meet the objectives of the New Zealand Public Health and Disability Act 2000, DHBs and the MOH are required to carry out principles-based prioritisation processes (MOH 2005). According to one Health Funding Authority report (HFA 2000), the principles that have consistently guided decision-making relating to publicly-funded health care in New Zealand are: effectiveness, equity, value for money and Maori health. In 2004 a Working Group composed of representatives from DHBs and the MOH developed a framework to assist decision-makers in prioritising health services*: "The Best Use of Available Resources: An approach to prioritisation"*. This framework, along with other tools such as the Health Equity Assessment Tool (a tool developed by MOH to tackle inequalities in health), is currently being used by many DHBs to guide purchasing decisions.[17] In addition, each year the Minister of Health (in conjunction with the MOH) sets out the national strategic priorities that are to be adopted by the health sector. For example, this year's health targets are: shorter stays in emergency departments, improved access to elective surgery, shorter waits for cancer treatment, increased immunisation rates, better help for smokers to quit, and improved diabetes and cardiovascular services.

As mentioned in Chapter 1, PHARMAC was created in 1993 as part of the New Zealand medicines system, to obtain the best possible health outcomes for New Zealanders within a defined budget.[18] PHARMAC decides on behalf of the DHBs which medicines will be subsidised for use in public hospitals and in the community. Prioritisation criteria are used within a programme budgeting and marginal analysis (PBMA) framework (explained in Section 2.4.3) to decide which medicines to fund. The criteria include cost-effectiveness, budgetary impact, the health needs of New Zealanders and in particular Maori and Pacific

[17] For example, the following criteria are included in the MidCentral DHB's prioritisation framework: 'alignment with DHB strategies', 'evidence to support proposal', 'equity summary', 'measurement, evaluation and reporting', 'procurement' and 'financial impact'.

[18] Earlier this year the Government confirmed that PHARMAC will extend its management of medicines to include hospital medical devices. It is estimated that the shift to full management will take five years (i.e. 2017).

peoples, availability and suitability of existing medicines and any 'other criteria' deemed to be relevant.[19]

2.2.6 Israel

In 1995 Israel passed the National Health Insurance Law which made health insurance compulsory, guaranteeing universal coverage of basic health care. The 'basket' of services provided by the largest existing 'sick fund'[20] at the time was used to establish a list of basic health services including treatments, medication and equipment to be provided to all residents (Shani et al. 2000). In 1997 there was a public outcry as the list of services had not been updated other than the addition of one drug. In 1999 The Medical Technologies Administration at the Ministry of Health was established to update and manage the national list of health services (NLHS) or as it is also known, 'the basket of health services'.

New technologies are evaluated by multi-disciplinary teams, based on clinical, epidemiological and economic factors, with clinical evaluation being the most important (Shani et al. 2000). This information is given to the Medical Technology Forum which grades new technologies on a scale from 1-10 using a set of guiding criteria including 'potential to prevent mortality or morbidity', 'number of patients to benefit', 'financial burden on society and/or patient' and 'net gain to society' (Sabik & Lie 2008). The rankings are then passed on to a National Advisory Committee (also known as 'the Basket Committee') made up of representatives from the government, the sick funds and the public. The Committee decides by consensus which technologies should be added to the list of basic health services with recommendations for their use, such as prescribing a drug according to clinical guidelines. The process has gained widespread acceptance in Israel "by government officials, healthcare professionals, politicians and the courts." (Shani et al. 2000, p 184)

---

[19] Hansen (2006) conducted a theoretical review of PHARMAC's over-arching approach to prioritising pharmaceuticals. The author recommended that PHARMAC be more explicit and transparent in its approach, to 'tighten up' their decision criteria and to consider whether the relative importance of the decision criteria should be determined in a more explicit way. It appears that PHARMAC has considered these recommendations. Some papers relating to the funding of a new-generation anticoagulant drug were recently released by PHARMAC. The papers reveal that PHARMAC's clinical advisory committee and cardiovascular subcommittee, considered this drug for nearly two years before PHARMAC agreed to fund the drug. During the process feedback was sought from a wide range of clinicians, pharmacists and patient groups. The papers also reveal that PHARMAC negotiated a substantial discount on the list price which made the drug more cost effective (www.pharmacy-today.co.nz (31/10/11)).

[20] A sick fund is a health insurance fund in which members pay fees to receive an agreed 'basket' of health services. Prior to 1995, enrolment in a sick fund was not compulsory.

This formal priority-setting process still exists today. Each year the Government determines how much is available to be spent on new technologies. Proposals from health plans, pharmaceutical companies, the Israel Medical Association, patient organisations and other groups are evaluated before the National Advisory Committee recommends what new technologies should be adopted (Rosen 2011).[21]

2.2.7 United Kingdom

In 1999 the National Institute of Clinical Excellence (NICE) was established as an independent health authority to provide guidance on health treatments and care for the health professionals, patients and public using the National Health Service (NHS) in England and Wales. NICE's three main functions are to appraise new technologies, to develop clinical guidelines and to provide guidance on public health (Raftery 2001).

Suggestions for technology appraisals come from a variety of sources including health professionals, patients, the general public, the National Horizon Scanning Centre and the Department of Health. When appraising health technologies, NICE relies heavily on cost effectiveness evidence (Devlin & Sussex 2011). The cost per Quality-Adjusted Life Year (QALY) [22] gained by a new treatment is compared against a comparable alternative treatment to determine cost effectiveness.[23] The new treatment's incremental cost effectiveness ratio (ICER) is compared against a threshold of £20,000-£30,000 to determine if it represents value for money. NICE also takes other factors into account when making judgements about cost effectiveness, such as severity of underlying illness, end of life treatments, stakeholder persuasion,[24] significant innovation, disadvantaged populations and age[25] (Rawlins et al. 2010).

---

[21] In 2011 the National Advisory Committee was presented with 430 technologies of which only 30% could be approved (The Jerusalem Post, 1 June 2011).

[22] A QALY combines length of life and quality of life into one measure. The number of QALYs gained from an intervention is calculated by multiplying the duration of survival by a utility weight representing a patient's health-related quality of life (HRQoL). For example, one year in perfect health is equal to 1.0 QALY, whereas one year in a health state with a HRQoL valued at 0.5 is equal to 0.5 QALY. Utility weights are calculated by using a number of different approaches including standard gamble, time-trade-off and the visual analogue scale.

[23] The additional cost of a new treatment is divided by the additional QALYs of the new treatment compared with a comparable alternative treatment, to produce an incremental cost effectiveness ratio (ICER).

[24] Stakeholders including patients, policymakers, purchasers, health professionals and scientists each have a different perspective in regard to priority-setting. It is now quite common for interest groups consisting of patients, their families and supporters to lobby politicians or funding agencies in an effort to secure funding for a new treatment. Stakeholder persuasion can influence funding decisions. For example, in 2008 PHARMAC agreed to fund full 12 month courses of the breast cancer drug Herceptin, a drug which had previously been funded for only nine-week courses, because of an election commitment made by the National party.

In 2002 an advisory committee was set up by NICE to incorporate the views of the public in NICE's decision-making. The Citizens' Council, made up of 30 members drawn from the population of England and Wales, discuss and report on the social, ethical and moral issues relating to health care prioritisation that might influence the development of NICE's guidance. NICE also consults widely to develop and disseminate its guidelines and priorities.

Although NICE has a consistent national-level approach to assessing evidence, sub-nationally the NHS is inconsistent in the way it makes its decisions (Mullen 2004, Devlin & Sussex 2011). Examination of priority-setting by health authorities within the NHS indicates that less weight is placed on cost per QALY compared with NICE's approach and that allocation decisions are largely being made on considerations other than cost per QALY (Appleby et al. 2009, Devlin & Sussex 2011).

2.2.8 Summary

In this section the explicit priority-setting approaches of a number of countries and Oregon, have been outlined. Two main theoretical approaches are apparent: an 'institutional' approach whereby principles guide prioritisation processes, and a 'technical' approach whereby countries have established committees or organisations with the purpose of establishing what services should be provided within a publicly-funded health system (Logan et al. 2004, Sabik & Lie 2008). Norway, the Netherlands, Sweden and Denmark tend towards an 'institutional' approach whereby principles such as 'human dignity' and 'solidarity' are used to guide priority-setting. On the other hand, Israel, New Zealand, the UK and the state of Oregon tend towards a more 'technical' approach. For example, New Zealand has instigated a points system in various specialities for prioritising patients and NICE uses a cost per QALY threshold for determining whether or not a service is to be funded. A technical approach allows for more consistency in decision-making as prioritisation is made explicit.

Assessing *new* health technologies within a priority-setting framework is discussed in the next section.

[25] The complexity of assessing the quality of life in children is taken into account when evaluating the cost effectiveness of treatments for children.

2.3 Health technology assessment and priority-setting

An increase in demand for new technologies has led many countries to establish separate agencies for the purpose of assessing new health technologies (Jonsson & Banta 1999). Evaluating new health technologies[26] before they become part of clinical practice is known as 'health technology assessment' (HTA) (Walley et al. 1998). HTA provides clinicians, managers and policy makers with information relating to a new technology to assist them in their decision-making. Traditionally HTA has focussed on the efficacy, effectiveness and cost-effectiveness of a new technology; however, other considerations including social, legal, ethical and political factors are now commonly included in HTA (AMGEN 2011).

HTA is typically undertaken by multi-disciplinary committees consisting of representatives from the government, insurance funds, health care providers, academics, health professionals, patients and the general public (Stafinski et al. 2011). In a review of HTA processes around the world, Stafinski et al. found there were three criteria common to most advisory committees: clinical need (encompassing severity, burden of illness, availability of alternative treatments), health impact (encompassing safety, efficacy and effectiveness compared with current care) and affordability (encompassing budget impact, number of patients and duration of treatment). Most committees also considered 'value for money' with many committees (e.g. NICE) using ICER thresholds. Criteria relating to 'social and equity' considerations were less common and information relating to ethical decisions was limited.

The use of HTA – at the national, local and institutional levels – to inform health care priority-setting is widespread (Stafinski et al. 2011). However, according to a review undertaken by Neumann (2009), the processes used, the participants involved and the level of transparency varies greatly between HTA processes both within and across countries. Whereas proponents consider that HTA promotes the efficient allocation of health care resources, opponents argue that HTA has been used to restrict access to new health care technology (O'Donnell et al. 2009). To ensure that HTA is successfully integrated into decision-making and accepted by the key stakeholders, O'Donnell et al. (2009) recommends that the HTA process is transparent and that key stakeholders are included in the process.

[26] New technologies include pharmaceuticals, devices, diagnostic tests, and medical and surgical procedures.

2.4 Establishing a prioritisation framework

There are a range of methods that are used to establish explicit priority-setting of health care including cost-utility analysis (CUA), for example league tables of QALYs; programme budgeting and marginal analysis (PBMA); contingent valuation methods such as willingness-to-pay (WTP); and multi-criteria decision analysis (MCDA) which encompasses choice-based methods such as conjoint analysis (CA) and discrete choice analysis (DCE). The methods should not necessarily be considered as 'stand-alone' tools for prioritising health care, as multiple and over-lapping methods can be used to establish a priority-setting framework (Peacock et al. 2009); for example, 'QALYs gained by treatment' can be a criterion in MCDA. These methods are discussed in the following sections.

2.4.1 Cost-utility analysis

A favoured analytic technique for economic evaluation in health care is CUA (Neumann et al. 2000). A common application of CUA is to use QALYs as a 'common currency' by which one health treatment can be compared with another in terms of the cost per QALY or the cost per QALY can be compared against a 'threshold'. For example, the predominant approach taken by NICE to prioritise health care includes four steps: calculate the QALYs per treatment; compare the cost per QALY with a set threshold; if the cost per QALY is below the threshold the treatment is considered to be cost-effective; if the cost per QALY is above the threshold, a treatment will only be provided if the additional cost can be justified.

However, the use of QALYs and a cost per QALY threshold has its limitations (Bryan et al. 2002, Dolan 2005). The QALY algorithm, as it is commonly used, assumes constant returns to scale[27] and does not capture additional considerations that may be important to society (Nord 1995, Dolan 1998, Devlin & Sussex 2011). Also, for some of the new health treatments being considered there may be limited information available in terms of efficacy, effectiveness or cost for instance, making it difficult to compare these new treatments with other treatments where QALYs are available (Raftery 2001). In addition, the threshold which is used to compare the cost per QALY for various treatments is often arbitrarily assigned (Detsky & Laupacis 2007) and when treatments are considered to be cost-effective, that is, when they fall below the threshold, the overall budget implication of funding that treatment is

[27] 'Constant returns to scale', in this context, assumes that the average cost and health effects (i.e. the number of QALYS gained) is independent of the number of patients treated. In addition, it is often assumed that costs are divisible (for example, if it costs $x to treat 10 patients then it is assumed it will cost $x/2 to treat five patients) (Reilly 2010).

not always considered (Devlin & Sussex 2011). Further, funding treatments based on 'value for money' does not take into account the opportunity cost of purchasing one health service/programme in place of another (Grocott 2009).

Rhodes et al. (2002) explored how additional considerations such as 'equity' could be incorporated into cost-utility analysis and suggested that when cost per QALY is greater than the threshold, consideration of six ethical issues can help determine how much 'weight' the cost per QALY should have in the decision to provide treatment. These six ethical issues are posed as the following questions. Should more money be spent per QALY on younger people than older people? Should more money be spent on saving the lives of 'identifiable patients' than saving the lives of 'statistical patients'? Should greater priority be given to those who are dying than a QALY calculation would suggest? When the gain in QALYs is the same for two patients, should the sickest patient receive preference? When there are no alternative treatments available should a new treatment be funded? When co-morbidities exist, the increase in QALYs from treatment might not be as high as expected – should allowances be made in these circumstances? The authors discuss these questions but do not provide answers. Instead they suggest that these ethical principles should be considered when rationing health care services.

Internationally, HTA bodies do take into account additional considerations when prioritising health care (Golan et al. 2010). For example, NICE applies 'special weightings' to specific factors, such as 'severity of underlying illness', 'end of life treatments' and 'disadvantaged populations' in addition to cost effectiveness (Rawlins et al. 2009).

However, taking into account additional considerations when evaluating health treatments is often done by way of qualitative judgements, which given the complex nature of decision-making, can lead to inconsistencies in the way decisions are made (Devlin & Sussex 2011). Devlin & Sussex argue that a more consistent and systematic approach is required to ensure accountability and transparency.

2.4.2 Programme budgeting and marginal analysis

Programme budgeting and marginal analysis (PBMA) provides a framework to assist decision-makers in allocating limited resources across a range of health services by using the concepts of opportunity cost and marginal analysis. When the budget is fixed the only way to

fund new treatments is to reallocate funding. To do this the marginal costs and marginal benefits of new treatments and/or programmes are considered alongside how resources are currently spent (Mitton & Donaldson 2003, Ruta 2005, Peacock 2006). The combination of technologies providing the greatest health gain within an available budget is selected, which may result in 'disinvestment' where resources are shifted away from currently funded technologies (Gallego et al. 2010). The benefit of using PBMA is that the opportunity cost of investing in a new technology is explicit (Grocott 2009). In New Zealand, The Pharmaceutical Management Agency (PHARMAC) uses a PBMA approach within a capped budget to prioritise technologies, mostly relating to pharmaceutical expenditure.

As discussed in the previous section, it is important that decisions involving multiple criteria are made in a consistent and transparent way. However, transparency of funding decisions may not always be possible or welcomed by the funding body. For example, negotiating within a capped budget encourages price competition – funding bodies can negotiate reduced prices and/or offer confidential rebates. For instance, PHARMAC uses "competitive processes such as tendering for supply" to encourage price competition (Grocott 2009 p 184). Also multiproduct agreements can be used to generate savings by purchasing a number of products from the same supplier (Grocott 2009). The end result is that the overall cost per QALY might fall within budget, but individual products might not have met the cost-effectiveness criterion had they been purchased separately.

2.4.3 Revealed preference and stated preference methods

As will be explained in Section 2.5.1, it is important to include the preferences of key stakeholders in the prioritisation process, particularly when health care is publicly-funded. Preferences can be measured in two ways – with stated preference or revealed preference methods.[28] As can be seen in Figure 2.2, Contingent valuation (CV) and MCDA are stated preference methods. Before describing CV and MCDA, I will explain what revealed preference and stated preference methods are and discuss some of the practical issues involved in selecting which method to use.

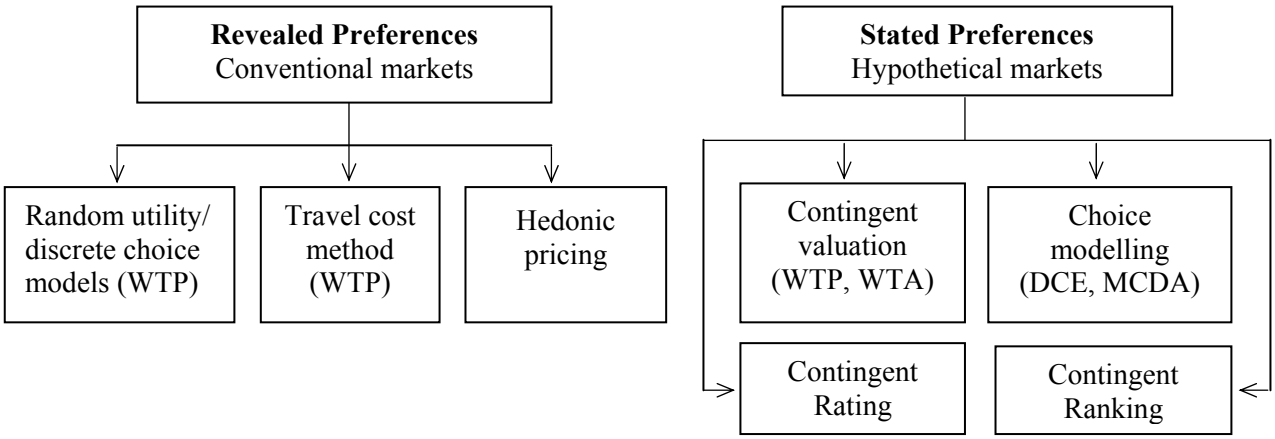
Revealed preference and stated preference methods are grounded in consumer theory – rational decision-makers seek to maximise their total utility (preferences) subject to a

[28] In order to derive utility weights for QALYs, the preferences of individuals for certain health states are elicited by techniques such as standard gamble, time-tradeoff, rating scales and ratio scales (Neumann et al. 2000).

constraint (Amaya-Amaya et al. 2008). With revealed preference methods individuals' preferences are revealed *indirectly* through the choices they make in markets (Samuelson 1948). By observing individuals' actual market behaviour the importance of attributes or variables affecting a decision can be ascertained. Revealed preference methods include discrete choice, travel cost and hedonic pricing. Discrete choice models assume that an individual has a number of alternative options and that the option they choose is the one that has the most preferred combination of attributes (Accent 2010). The travel cost approach, initially developed to value recreational assets, uses the cost of travel as a measure of preference for that activity. With hedonic pricing the value of an attribute is its *implicit* price or *shadow* price. For example, increased occupational risks may be reflected in higher wages or the value of a 'statistical life' can be inferred by the amount individuals are willing to pay for life-saving products such as smoke alarms or airbags for cars.

In contrast, with stated preference methods, the preferences of key stakeholders are *directly* revealed by the decisions they make in a contrived framework; for example, rating, ranking or choosing between hypothetical scenarios, using contingent valuation methods such as willingness-to-pay (WTP) or willingness-to-accept (WTA), and choice modelling (Bridges et al 2003). (These methods are explained in more detail below.)

Figure 2.2: Revealed preference and stated preference methods



With revealed preferences methods the choices that are observed are real market choices (i.e. individuals actually spend money, time or other resources). Therefore the estimated preferences using revealed preference methods are likely to be more accurate than the preferences obtained using stated preference methods where choices are contrived and

individuals do not actually spend money, time or other resources (Accent 2010). However, as health care is not traded in markets in the same way as other goods, a revealed preference method can lead to inaccurate estimates of preferences. For example, publicly-provided health care or health care subsidised by insurance companies results in many health treatments being free or low-priced and therefore the prices individuals pay for health care are unlikely to reflect their true willingness-to-pay. In addition, because of asymmetric information (i.e. doctors or health professionals often have more information than the patient) the decisions made by patients may not be based on their actual preferences (Ryan et al 2008). Further, with revealed preference methods, preferences for 'non-use' goods (e.g. new health services which are yet to reach the market) cannot be estimated, something that is possible with stated preference methods.

The use of a *joint* revealed preference and stated preference method is becoming popular (Cherchi & Ortuzar 2006). It allows researchers to combine data based on actual observations with data obtained from contrived settings (i.e. when no actual observations are available). If actual observations are available, revealed preference data can be used as a tool to cross-validate stated preference data.

The two main categories of stated preference methods are CVM and choice modelling techniques (i.e. MCDA). These methods are discussed in the following sections.

2.4.4 Contingent valuation methods

Contingent valuation methods (CVM) include willingness-to-pay (WTP) and willingness-to-accept (WTA). CVM provide estimates of the value of health treatments or programmes by asking individuals hypothetical questions relating to how much they would be 'willing-to-pay' for a particular health treatment and/or programme (WTP) or how much they would be 'willing-to-accept' in compensation to give up, or lose access to, a health treatment and/or programme (WTA).

CVM can be targeted to specific population groups, depending on the objective of the study. For example, the general population could be surveyed to value insurance premiums for specific programmes, or users of a particular health programme could be surveyed to value the associated programme costs (Bavoumi 2004). CVM can be used to value any aspect of

health care and is not restricted to particular considerations such as the length of life and HRQoL, a criticism associated with the use of QALYs[29] (Olsen 1997).

However, CVM have been widely criticised in the literature in regard to their inaccuracy and inconsistency in valuing health care (Smith 2003, Venkatachalam 2004). Individuals may understate or, alternatively, exaggerate their willingness-to-pay either deliberately or unintentionally. For instance, individuals may find it difficult to attach a value to a health treatment, particularly when they are unfamiliar with the treatment, or they may have a moral objection to doing so (Gerard et al. 2008). Opponents of CVM argue that using either approach – WTP or WTA – to value a treatment or programme, should result in a similar value (Ahlheim & Buchholz 2000). However, it has been demonstrated in the literature that there are substantial disparities between WTP and WTA when valuing the same issue (Mitchell & Carson 1989), leading to concerns about how well CVM measure individuals' preferences (Ahlheim & Buchholz 2000, Venkatachalam 2004).

According to Gerard et al. (2008), the approach used to value health care or elicit preferences, depends on the level of information required. CVM is considered suitable for valuing overall programmes but when more information is required on the various characteristics that make up a programme, then other approaches such as discrete choice experiments[30] are more useful.

2.4.5 Multi-criteria decision analysis

Priority-setting of health care involves making trade-offs between multiple conflicting considerations or 'criteria' (Mullen 2004). Formal multi-criteria decision analysis (MCDA) approaches, encompassing a range of techniques and methods, have been developed in many areas including health care, to assist individuals or groups in making complex decisions involving multiple criteria in an explicit, consistent and transparent way (Belton & Stewart 2002). What MCDA approaches have in common is identifying the relevant criteria to be considered and estimating the relative influence these multiple criteria have in the decision-making process (Devlin & Sussex 2011).

[29] A criticism associated with the use of QALYs is that emphasis is placed on length of life and quality of life and though other considerations may be taken into account when prioritising health services, it is often not done explicitly. This is discussed further in the next section.
[30] Discrete choice experiments are discussed in Section 2.7.2.

MCDA or MCDA-based approaches are becoming increasing common in health care prioritisation. A *typical* MCDA process includes the following elements (Baltussen 2006, Bridges 2011):

1. Identify the decision context – what are the alternatives (health treatments in this case) to be prioritised and who should be involved in the process of identifying and establishing the relative importance of the criteria.

2. Identify the relevant criteria by which the alternatives will be ranked including all possible considerations such as strategic, social, economic and ethical considerations.

3. Determine the relative importance of the criteria – that is, decide how trade-offs will be made across the criteria to obtain 'weights' for each criterion.

4. Use the criteria weights to assist in decision-making – for example, score the alternatives using the criteria weights to obtain a ranking of the alternatives.

In addition, performing a sensitivity analysis – for example, exploring whether different weights affect the overall ranking of the alternatives – often becomes an important part of the prioritisation process (Devlin & Sussex 2011). There also needs to be scope for "ongoing reflection and review" throughout the process (Belton & Stewart 2002, p 38) – criteria, weights and/or alternatives might need to be revised when new information becomes available for instance.

MCDA complements other priority-setting approaches such as PBMA and CUA (Devlin & Sussex 2011). For example, as mentioned earlier, 'QALYs gained by treatment' could be included as a criterion in a MCDA prioritisation framework alongside other criteria.[31] Because of its transparent and systematic approach, and given that it incorporates the preferences of key stakeholders in the decision-making process, a MCDA approach is used in this thesis to establish a prioritisation framework.

The elements of a typical MCDA process, as outlined above, are discussed in the following sections.

[31] This is discussed in Chapter 5.

2.5 Determining the relevant criteria

The main objective of this thesis is to establish a prioritisation framework that aids decision-making predominantly at the meso-level of health care (i.e. allocating a fixed health budget across a range of health and disability services). The decision criteria, by which alternative health treatments and/or programmes will be prioritised, need to be determined. It is a crucial stage of the process – the criteria need to be relevant to the decision being made, ideally they need to be independent of each other, and the alternatives being considered need to be accurately described on the criteria, otherwise the overall ranking of alternatives may not be accurate. Deciding *who* should be involved in determining the criteria and its relative importance is an important part of the process.

2.5.1 Whose views?

There appears to be a lack of clarity and consensus in the literature about what it means for the public to be involved in health care, why such involvement is desirable, how it should be achieved and how it is incorporated by decision-makers into priority-setting processes (Florin & Dixon 2004, Mitton et al. 2009).

The public can be involved in priority-setting in two main ways: by 'informing' prioritisation decision-making or by taking an 'active' role in setting priorities and allocating resources.[32] The first approach uses the values and principles of key stakeholders to 'inform' prioritisation decision-making. The second approach allows key stakeholders to make priority-setting decisions. A criticism of the 'active' approach is that some prioritisation decisions require significant clinical knowledge and members of the public may lack the ability to make appropriate decisions (Lenaghan 1999, Florin & Dixon 2004, Logan 2004).

Florin & Dixon (2004) argues that the general public should be involved in priority-setting for two reasons. First, public health services are indirectly financed by the public and therefore the public should be involved in deciding what services are to be funded, and second, by involving individuals and communities in priority-setting, health services will be tailored to specific needs which results in improved health outcomes. Sampietro-Colom et al. (2008) holds a similar view and suggests that under a tax-based system, the participation of all

[32] In contrast, 'patient' involvement means that patients are involved in decision-making at the treatment level; for instance, assisting in the development of clinical trials or reviewing particular health services (Boote et al 2010).

affected individuals – consultants, general practitioners, nurses and allied-health professionals, social workers, patients and their relatives and the general public – is preferable as the values and preferences of the various sub-groups may differ.

Bruni et al. (2008) regards members of the public as the most important stakeholders in the health care system and that as part of a democratic process their values and priorities need to be considered. The authors suggest that engaging the public leads to better decision-making and increases the confidence the public has in the health system. Lenaghan (1999) agrees. The author suggests that within a democracy, citizens have a right to be involved in decisions that affect them. Involving the public in explicit rationing decisions is beneficial because the public becomes aware of the need to ration health services due to limited funding, and being part of the process encourages greater confidence in the health sector. However, as Lenaghan (1999) points out, involving the public in priority-setting needs to be done for the value it adds to the process and not merely as a token gesture.

Peacock et al. (2006) advocates community involvement in priority-setting and suggests that good representation is vital to ensure that decisions reflect the values and needs of the community. In addition, the authors recommend that the priority-setting process, as well as any decisions made, is accessible to managers, doctors, patients and the public.

Wiseman et al. (2003) surveyed 373 Australian citizens about whether the general public has a legitimate role in informing priority-setting of health care. The authors found there was strong support for the 'general public' to be involved in priority-setting at all levels of health care, particularly across broad health care programmes and in deciding what criteria should be used to allocate funds across different population groups. On the other hand, there was strong support for medical professionals and health service managers to prioritise health treatments and medical procedures. However, most participants agreed that the preferences of a range of groups need to be included in the decision-making process.

In Litva et al.'s (2002) study, a mix of focus groups and in-depth interviews were used to explore the views of the general public in the UK with respect to public involvement in health care decisions. Respondents included randomly selected members of the public and workers from health and non-health related organisations. The authors found that the respondents were willing to be consulted about decisions relating to the 'system and programme' levels as long as their input was valued and they were not held accountable for the final decisions. However,

they were less willing to be involved in decisions relating to individual patients and felt that the only involvement the public should have at this level was in establishing criteria for prioritising patients.

A range of methods can be used to elicit the views of key stakeholders on priority-setting, including public consultation, questionnaires, focus groups, citizens' juries and surveys.[33] The choice of method needs to be suited to the purpose of the study and the sample should be representative of the population affected (Florin & Dixon 2004, Belton & Stewart 2002). In addition, by ensuring wide societal participation from the beginning of the priority-setting process, greater acceptance and trust in the outcome will follow (Whitehead 1991, Lenaghan 1999, Bruni 2008).

In my thesis key stakeholders are involved at every stage of the process: health professionals assisted in the development of 14 health vignettes (Chapter 3); six focus groups comprising members of the general public, medical and non-medical health care workers, public health workers, a Maori health provider and a GP practice suggested potential prioritisation criteria (Chapter 4); and to find out how important the criteria are, three groups of respondents completed an online decision survey[34] including a randomly selected sample of the New Zealand public, a non-randomly selected sample and a sample of health economists (Chapters 5 and 6).

2.6 Determining the relative importance of the criteria and their use in decision-making

After the appropriate decision criteria have been identified, the relative importance of each criterion in the decision-making process needs to be estimated. Baltussen & Niessen (2006) explains how a performance matrix, which is common to most MCDA, can be analysed qualitatively and/or quantitatively. A performance matrix breaks down the alternatives (health treatments and/or programmes) to be assessed on the criteria on which they will be measured. For example, Table 2.1 displays a hypothetical performance matrix for two treatments 'vaccine for cervical cancer' and 'hip replacements' which are described on four criteria.[35]

[33] These methods are discussed in Chapter 3.
[34] In the online decision survey respondents are asked to choose between two pairs of criteria, thereby trading-off one criterion for another.
[35] A performance matrix can include any number of alternatives and/or any number of criteria/levels.

Table 2.1: Performance matrix

Alternative	Cost per QALY	Benefit to society	Alternative treatment available	Age
Vaccine for cervical cancer	$4210	Low	Yes	14-18 years
Hip replacements	$5151	Medium	No	55 years and over

To assess the performance of the alternatives based on the criteria, the criteria need to be assigned levels. The levels can be categorical (yes or no), ordinal (low, medium, high) or cardinal (a cost per QALY of $600 is twice that of a cost per QALY of $300) (Ryan 2000). Decision-makers can qualitatively rank the alternatives by determining whether any alternative 'dominates' by performing better than the other alternatives on at least one criterion and at least as well on the other criteria or when no alternative 'dominates', by making subjective judgements. Qualitative analysis is quick and might help in decision-making but decisions based largely on subjective reasoning are unlikely to be consistent and could result in an undesirable ranking of alternatives particularly when no alternative 'dominates'.

With quantitative analysis, the information in the performance matrix is converted into numerical values using various MCDA techniques (Baltussen 2006). Most of the quantitative MCDA techniques are based on multi-attribute utility theory (MAUT)[36] or outranking methods (Fülöp 2005). Several methods based on MAUT, and the outranking method will be discussed in the next section.

The MAUT method is known as a 'compensatory method' as high scores on one criterion can compensate for low scores on another (Keeney & Raiffa 1976).The most common MAUT method for aggregating criteria weights is the simple additive multi-attribute value model (Mullen 2004). The models are also known as 'linear', 'scoring' or 'point-count' systems or models .The total value of an alternative (health treatment and/or programme) is calculated by multiplying the value score on each criterion by the weight of the criterion and then adding the weighted scores together.

[36] The basis of MAUT or multi-attribute value theory (MAVT) is that any good or service (in this case health treatments and/or programmes) can be described by their characteristics (attributes/criteria). The extent to which a treatment is valued depends on the utility gained from each of the attributes/criteria (Lancaster 1966). 'Attributes' is synonymous with 'criteria'. The term 'attributes' is commonly used in conjoint analysis and discrete choice experiments.

A simple linear additive model takes the following form (Mullen 2004, p 50):

$$P_j = \Sigma\, w_i \,.s_{ij}$$

where P_j is the priority score of the jth alternative treatment, s_{ij} is the score or rating of the jth treatment on the ith criterion and w_i is the weight or value of the ith criterion.

Treatments can then be ranked based on the overall scores. A simple additive model is applicable only if the criteria can be 'traded-off' and the criteria are independent of each other. This means that the preferences for two or more criteria are independent of the fixed levels of the remaining factors (i.e. there are no interactions across the criteria) (Krantz 1972). For example if Alternative A has three factors, a,b,c and Alternative B has three factors, a'b'c', and if a is ranked ahead of a', b is ranked ahead of b' and c, c' are tied in their ranking, then Alternative A is ranked ahead of Alternative B, independent of the tied ranking of c,c'.

The use of a simple additive model is illustrated in the following example. Wilson et al. (2006) conducted a one-day workshop to develop a prioritisation framework for use in an English Primary Care Trust (PCT).[37] Twenty representatives from across the local economy took part. Participants were divided into five groups to 'brainstorm' the relevant criteria. Criteria weights were established by the participants allocating 100 percentage points amongst the criteria; for example the mean weight (for the five groups) was 19.8% for 'need', 18.2% for 'quality of life' and so on. Four hypothetical programmes were then scored against each criterion on a scale of 0-10. For each programme, the score on each criterion was multiplied by the corresponding weight and then summed across the criteria to obtain a weighted benefit score. Finally the programmes were ranked in order of their cost-value (the weighted benefit score divided by cost).

Several MCDA methods that use a simple additive model are discussed in Sections 2.7.1 to 2.7.4 and the outranking method is discussed in Section 2.7.5.

[37] PCTs are part of the English National Health Service (NHS). They are responsible for commissioning health services within an allocated budget for the population under their authority (Wilson et al 2006).

2.7 Methods for establishing criteria weights

There are many formal methods for eliciting preferences and aggregating weights. In this section several of the main methods are discussed. For a detailed analysis of MCDA methods, see Belton & Stewart (2002).

Criteria weights can be determined directly or indirectly. Two traditional methods that use direct rating techniques and ranking to explicitly define criteria weights include SMART (Simple Multi-Attribute Rating Technique) and SWING. With choice modelling criteria weights are determined indirectly. Decision-makers are presented with two or more hypothetical options consisting of a number of criteria that vary over a range of levels and asked to either rank, rate or choose between the options[38] and then the particular algorithm determines the weights corresponding to rankings, ratings or choices. Methods that use choice modelling to determine criteria weights include DCEs (Discrete Choice Experiments), AHP (Analytical Hierarchy Process) and PAPRIKA (Potentially All Pairwise RanKings of all possible Alternatives). These methods are discussed in this section.

2.7.1 SMART (Simple Multi-Attribute Rating Technique) and SWING

SMART (Edwards 1977, Edwards & Barron 1994) and SWING (von Winterfeldt & Edwards 1986) are simple multiattribute weighting methods based on ratio estimation.[39]

With SMART, weights are elicited in two stages. First, decision-makers rank the criteria in terms of their importance from the most important criterion for the problem at hand, to the least important criterion. Second, a fixed number of points (usually 10), is assigned to the least important criterion. Then more than 10 points are assigned to the second least important criterion to reflect its relative importance and so on until all the criteria have been assigned progressively higher points. The criteria scores are then normalised to one to obtain the final weights.[40]

SWING weighting takes into account the criteria levels when estimating criteria weights. Usually this involves the decision-maker being presented with a 'worst case scenario' (that is,

[38] Maximum difference scaling is another method that can be used where respondents

[39] These methods are based on the Expectancy-Value Theory proposed by Fishbein (1963). 'Maximum difference scaling' is another method based on ratio estimation where respondents are presented with choice sets and asked to rate the most important and least important attributes (Flynn et al 2007).

[40] The scores are normalised by dividing the points assigned to each criterion by the total number of allocated points.

a hypothetical alternative consisting of all the criteria at their worst levels) and then he or she is asked to identify the most important criterion which they think should be moved from the worst level to the best level. For example, consider a worst case scenario of 'living in constant pain', 'unable to work' and 'unable to care for oneself'. The decision-maker is asked to identify the criterion that they would prefer to move to the best level, i.e. 'living without pain', 'able to work' and 'can care for oneself'. If the decision-maker considers that moving from 'living in constant pain' to 'living without pain' is more important than the swing from 'unable to work' to 'able to work' and 'unable to care for oneself' to 'can care for oneself' then they assign 100 points to the 'pain' criterion. The decision-maker is then asked to select another criterion to be moved from the worst level to the best level and to assign points less than 100 to this change. This process continues with the remaining criteria (in this example, there are only three criteria). The assigned points are normalised to obtain the final weights. Alternatives can then be ranked according to their weighted scores.

Since it was originally developed SMART has been adapted to SMARTS to include SWING weighting (Edwards & Barron 1994). SMART has also been extended to reduce the amount of input required by the decision-maker (SMARTER) (Edwards & Barron 1994).

2.7.2 DCEs (Discrete Choice Experiments)

DCEs (or CA) [41] elicit the preferences of decision-makers for different aspects of healthcare by asking them to choose between two or more choice sets (Amaya-Amaya et al. 2008). Decision-makers may also be given the option of 'neither'. By analysing the choices decision-makers make, attribute[42] weights that reflect the relative importance of each attribute are estimated statistically (Ryan & Gerard 2003). An example of a full-profile trade-off question is presented in Figure 2.3.[43]

[41] Many researchers use the terms conjoint analysis and DCEs interchangeably. However, Louviere et al (2010) argues that there is a clear distinction between the two. The authors contend that unlike conjoint analysis, DCEs are modelled within a random utility maximisation framework (McFadden 1974). Random utility theory (RUT) is based on the theory of decision-making and choice behaviour and separates the preferences of respondents into two components: a systematic component (which measures the utility of the attributes) and a random component (which encompasses unobservable factors that may influence choice and measurement errors) (Manski 1977). For a full discussion about the distinction between conjoint analysis and DCE, see Louviere et al (2010).

[42] Typically with discrete choice experiments and conjoint analysis, the term 'attribute' is used instead of 'criterion'.

[43] A 'full-profile' choice set includes all criteria at differing levels whereas a partial-profile choice set includes a sub-set of criteria at differing levels.

Figure 2.3: Example of a full profile trade-off question from a DCE

Attribute	Hypothetical interventions	
	A	B
Severity of disease	Severe	Not severe
Number of potential beneficiaries	Small	Large
Age of target group	Young	Elderly
Individual health benefits	Small	Large
Poverty reduction	Neutral	Positive
Cost-effectiveness	Not cost-effective	Cost-effective
Which one would you choose? (Please tick a box)	☐	☐

Source: Baltussen & Niessen (2006)

The total number of choice sets presented to decision-makers depends on the number of attributes and levels and the experimental design. As the number of combinations of attributes and levels increases the number of potential profiles increases exponentially. Fractional factorial design is often used, where a subset of all possible combinations of attributes are selected, to limit the number of choice sets presented to decision-makers in order to reduce information overload and elicitation burden (Baltussen et al. 2007). However, care needs to be taken to ensure that the number of choice sets used (combining the various attributes and levels) results in enough data for statistical analysis.

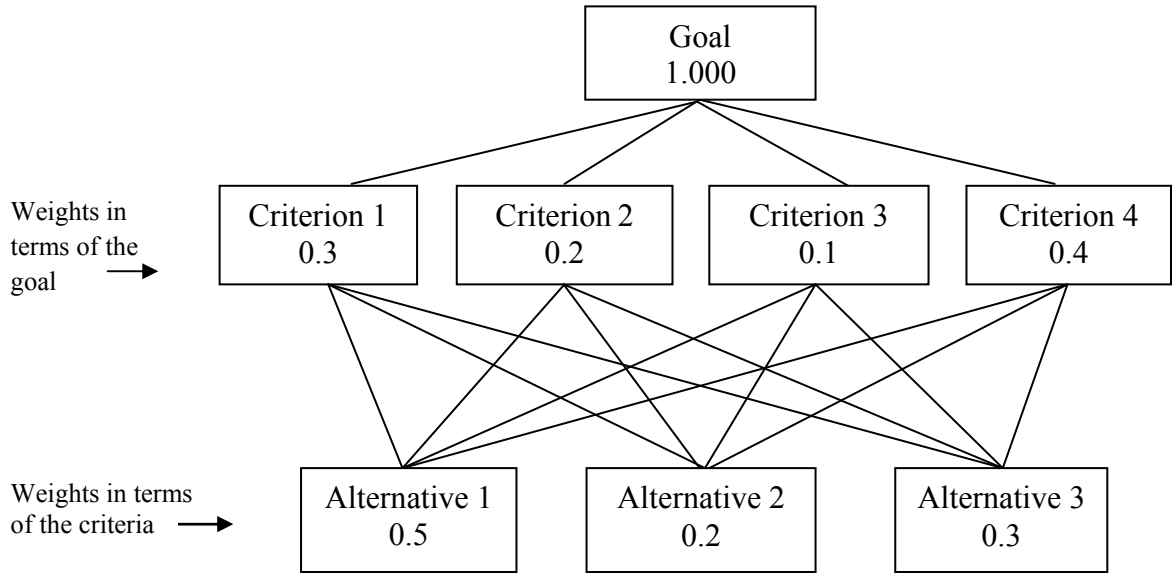
Adaptive conjoint analysis (ACA) is considered a 'hybrid' of conjoint analysis as it also allows 'self-explication' where decision-makers rate or rank particular attributes before answering pairwise trade-off questions. ACA, implemented through Sawtooth Software, enables a greater of number of attributes and levels to be considered compared with traditional CA methods (Johnson 2001).

Various estimation procedures such as probit, logit and multinomial logit are used to produce a set of weights for the sample (Ryan & Gerard 2003, de Bekker-Grob et al. 2012*)*. Each weight reflects the relative importance of an attribute averaged across respondents, the willingness to trade one attribute for another and willingness-to-pay if cost is included as a criterion. Interaction terms in respondent characteristics can be included in the statistical model to estimate how weights vary, on average, with those characteristics. The weights can also be used to derive total benefit scores which can be used to rank treatments and/or services (Ryan 1999).

2.7.3 The Analytic Hierarchy Process (AHP)[44]

AHP, as the name implies, is a hierarchical process where decision problems are decomposed into hierarchies as illustrated in Figure 2.4. Pairwise comparisons of the elements at each hierarchical level provide the data to estimate the weights for the criteria and/or alternatives in terms of the overall goal.

Figure 2.4: Simple AHP hierarchical process



Source: Adapted from Wikipedia (weights have been added for illustrative purposes)

For instance, in Figure 2.4 the weight for an alternative is estimated by asking decision-makers to state how much more important one alternative is compared to another in terms of a criterion (represented by the lines from each criterion going to each of the three alternatives). That is, in terms of *each* criterion, how much more important is Alternative 1 compared with Alternative 2; how much more important is Alternative 1 compared with Alternative 3 and how much more important is Alternative 2 compared with Alternative 3. Decision-makers answer using a nine-point intensity scale. The scale is displayed in Figure 2.5.

[44] For a full description of AHP, see Saaty (1994).

Figure 2.5: AHP nine-point intensity scale

Intensity of importance	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance
2,4,6,8	Compromises between the levels
1.1, 1.2 etc	For criteria that are very close in importance
Reciprocals of above	In comparing criteria *i* and *j*, if *i* is 3 compared to *j*, then *j* is 1/3 compared to *i*

Source: Saaty (2008)

A matrix is then constructed using the *relative* values (preferences) for each of the alternatives.[45] The 'alternative' weights (based on each criterion) are determined by solving for the principal eigenvector.[46] To assess the consistency of a decision-maker's judgements, a consistency ratio is calculated. Although reciprocal judgements involving *one* pair of criteria will be consistent, inconsistent judgements can arise between pairs of criteria.[47] The alternative weights and consistency ratio can be calculated automatically by using AHP software such as *Expert Choice*.

The next step is to compare the performance of each criterion with respect to the overall goal (represented by the lines going from the goal to each of the criteria). The same process occurs where a matrix of the judgements is used to produce weights for the criteria.

An *overall* weighted score for each alternative can then be calculated. Each criterion weight is multiplied by the weight of the alternative being considered and summed to obtain an overall score in the same way as a multi-attribute value function (Belton & Stewart 2002). For example, the value of Alternative 1 is calculated by multiplying 0.5 by the weight for each criterion and adding them together. The overall score for Alternative 1 is (0.5 × 0.3) + (0.5 × 0.2) + (0.5 × 0.1) + (0.5 × 0.3) = 0.5. The overall scores for Alternatives 2 and 3 are 0.2 and

[45] To determine a set of relative priorities amongst the three alternatives (n) only n(n-1)/2 judgements are actually required, resulting in a 3 x 3 matrix. For example, if a respondent considers Alternative 1 to be four times as important as Alternative 2 in terms of Criterion1, then the relative values are 4, and ¼.

[46] The AHP method was originally developed by Thomas L. Saaty. For a full mathematical explanation of how the weights are derived from the matrix format see Saaty (1994).

[47] A consistency ratio (CR) is calculated by measuring the consistency of a respondent's judgements compared with a large sample of random judgements. If the CR is greater than 0.1 the respondent's judgements are considered to be unreliable.

0.2 respectively. Therefore, based on the overall score Alternative 1 is chosen. Cost can be included as a criterion in an AHP survey or it can be considered separately alongside the weighted scores.[48]

A problem associated with AHP is 'rank reversal'. This occurs when adding a new alternative or removing an alternative, reverses the ranking of existing alternatives. To address this problem, there are two versions of AHP: the 'ideal mode' where ranks are preserved when a new alternative is added or an alternative is removed and the 'distribute mode' which allows the ranks to change (Millet & Saaty 2000). Brugha (1998) argues that if the criteria are modelled correctly then adding or deleting alternatives should not be a problem. That is, if adding a new alternative reverses the ranking of existing alternatives, then it is likely that a criterion is missing from the model.

2.7.4 PAPRIKA[49]

The PAPRIKA (Potentially All Pairwise RanKings of all possible Alternatives) method is used in this thesis to establish criteria weights. The reasons why this method has been chosen over other methods will be explained in Section 2.8.

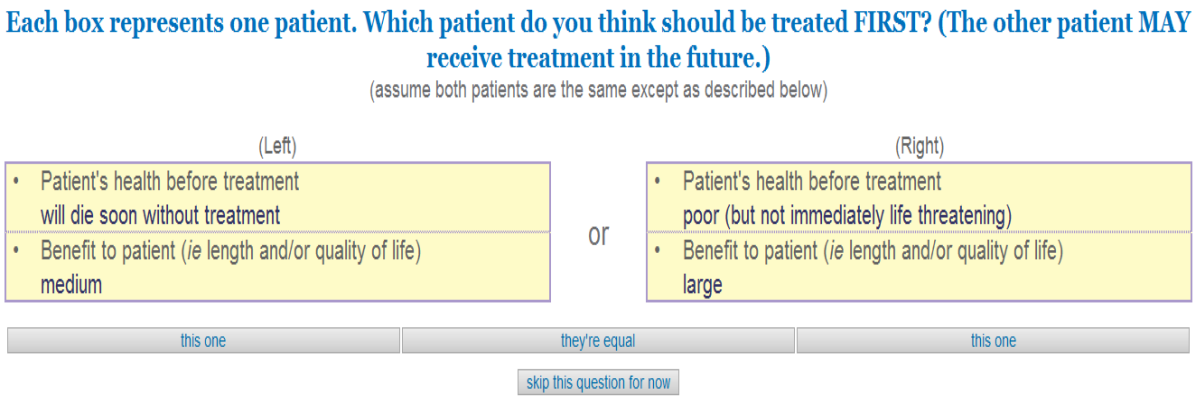
The PAPRIKA method is implemented through 1000Minds software [www.1000Minds.com] (Ombler & Hansen 2012). Decision-makers are presented with a series of hypothetical choices, each of which involves two imaginary patients who differ in only two characteristics.[50] Each choice requires the decision-maker to trade-off one characteristic for the other. Figure 2.6 is an example of a trade-off question from a 1000Minds online survey. Decision-makers express an ordinal preference by choosing which of the two imaginary patients should be treated first.

[48] If there are more than seven alternatives to be considered, an extra level can be added to the hierarchy (above the alternatives) which consists of rating categories for each criterion (e.g. high, medium, low). The weights for the categories are determined by making pairwise comparisons between the categories for each criterion. Alternatives are then evaluated using the appropriate rating category for each criterion.

[49] For an indepth explanation of the PAPRIKA method, see Hansen & Ombler (2009).

[50] As will be explained later, more than two characteristics can be included in each alternative but simulations of the PAPRIKA method reveal that this is not necessary to obtain an overall ranking of alternatives.

Figure 2.6: Example of a trade-off question using the PAPRIKA scoring method



The software automatically changes the order of the trade-off questions for each decision-maker; that is, the first question presented to one decision-maker is unlikely to be the same as the first question presented to another decision-maker. Changing the order of questions, on average, reduces or eliminates potential 'order biases'.[51]

Any number of criteria and/or levels can be included in the survey. However, as the number of criteria and levels increases the number of potential alternatives (combinations) increases exponentially. For example, with six criteria and four levels, there would be 4096 possible alternatives (4^6). To rank these from 1 to 4096, 8,386,560 pairwise combinations, comparing each alternative relative to every other alternative, would be required, i.e. (40962 – 4096)/2 = 8,386,560 (Hansen & Ombler 2008). The PAPRIKA method drastically reduces the number of choices that decision-makers have to make by automatically excluding 'dominant' pairwise comparisons and using the property of transitivity to implicitly answer other questions (which will be explained shortly). In addition, some of the combinations will be impossible. For example, 'benefit to patient: large' and 'patient's health before treatment: relatively good'. 'Impossible' combinations can be entered into the software so that decision-makers are not presented with unrealistic trade-off questions.

As mentioned above, the software automatically excludes 'dominant' pairwise comparisons. 'Domination' occurs when one alternative has a higher rating on at least one criterion and none lower on any other criterion, compared with the other alternative. This means that one alternative 'dominates' the other and therefore a decision is not required. For example, the combination of 'benefit to patient: large' and 'patient's health before treatment: will die soon'

[51] Order bias occurs when the sequence of survey questions influences the way a respondent answers a survey (Landon 1971, Perreault 1975, Dillman 1999).

dominates the combination of 'benefit to patient: small' and 'patient's health before treatment: will die soon'.

This leaves the 'undominated' pairs to be resolved. An 'undominated' pair is where one alternative has at least one criterion with a higher rating *and* at least one criterion with a lower ranking compared to the alternative. For example, the choice in Figure 2.6 above – the imaginary patient on the left has a higher rating on 'patient's health before treatment' but a lower rating on 'benefit to patient' compared with the imaginary patient on the right. A decision-maker is therefore required to make a choice.

When a decision-maker explicitly makes a choice between two 'undominated' pairs, the software automatically eliminates all other potential choices that are implicitly answered as corollaries of that choice via transitivity. Put simply, if patient A is ranked higher than patient B and patient B is ranked higher than Patient C, then by transitivity, patient A is ranked higher than patient C. After the first two choices have been made, the third choice is redundant. The software then automatically selects another choice randomly from the remaining 'undominated' pairs, and the process continues until all 'undominated' pairs have been ranked.

'Point values' or 'weights', which represent the relative utilities of each level of each criterion, are obtained by solving the inequalities (strict preference) and equalities (indifference) with a linear programme:[52]

Min $a_1 + ... + a_i + ... + n_1 + ... + n_j$	*a* through *g* are criteria, *i* and *j* are number of levels
s.t.:	
$a_1, ... n_j \in I^+$	All 'weights' are positive integers
$a_{i+1} - a_i, \geq 0 \; \forall$ a through n	Utility is positive monotonic
$a_1 + b_2 > = < b_1 + a_2$	Each explicit ranking

The PAPRIKA method can be implemented as a full or an incomplete ordinal information method. Using a full ordinal information method, a decision-maker starts by ranking 'undominated' pairs defined on just two criteria and continues to choose between pairs with progressively more criteria until *every* 'undominated' pair is ranked. All pairwise rankings of

[52] The utility function, which will be discussed in Chapter 7, can be expressed as follows. For individual i: $U_i = U_i [f_{ia}(a) + f(_{ib}(b) + ... + f_{ij}(j)]$ where a, b,...j are the criteria and $f(\cdot)$ are linear and positive monotonic in the levels of each criterion.

all possible alternatives have been identified (i.e. the explicitly ranked pairs and the pairwise rankings implied by them) providing an overall ranking of *all* possible alternatives for each decision-maker.

On the other hand, with an incomplete ordinal information method, decision-makers do not rank *every* 'undominated' pair resulting in an overall ranking that may or may not be close to a decision-maker's *true* overall ranking had they ranked *every* 'undominated' pair. Hansen & Ombler (2008) ran extensive simulations of PAPRIKA to test the accuracy of an overall ranking when a decision-maker does *not* rank every 'undominated' pair relative to a decision-maker's *true* overall ranking and found that ranking pairs defined on only two criteria is sufficient to produce an overall ranking which is highly correlated with the ranking that results from ranking every possible alternative.

Continuing with the example given previously, with six criteria and four levels, a decision-maker is faced with 8,386,560 choices. After the 'dominant' pairwise comparisons are excluded, 7,390,656 pairwise comparisons remain. To obtain a complete overall ranking 120 explicit pairwise decisions need to be made. However, given the results of the simulations as discussed above, approximately 60 decisions are sufficient to achieve an overall ranking that although incomplete, is highly correlated to the true overall ranking.

Hansen & Ombler (2008) found that the number of pairs decision-makers can comfortably rank in a short time is around 50 to 100 and that on average fewer than 100 explicitly ranked pairs (with two criteria) would need to be ranked for larger value models. (For example, a value model with four criteria and seven levels or eight criteria and four levels.)

1000Minds software (Ombler & Hansen 2012) and the PAPRIKA method (Hansen & Ombler 2008) have been used by researchers in many different areas including strategic management, the agricultural industry, commerce, environmental resources management and health related-fields.[53] 1000Minds has also been used to create points systems for prioritising patients for elective health services. Since 2004 point systems have been developed in New Zealand for coronary artery bypass graft surgery, hip and knee replacements, varicose veins surgery, cataract surgery, gynaecology, plastic surgery, otorhinolaryngology and heart valve surgery

[53] For example, Smith & Fennessy (2011) used1000Minds to determine the relative importance of specific traits associated with pasture species in Australia; and The American College of Rheumatology and the European League Against Rheumatism used 1000Minds to establish a new classification system for rheumatoid arthritis (Neogi et al 2010).

with further points systems planned for the future (Hansen et al. 2012). Similar priority-setting scoring systems are being used in Canada for services such as cataract surgery, general surgery procedures and hip and knee replacements (Noseworthy et al. 2003).

2.7.5 Outranking methods[54]

Unlike the methods described above, an outranking method does not impose an underlying aggregate value function and therefore alternatives are not ranked based on an overall value or score. Instead, an outranking method compares each alternative in terms of the criteria to obtain pairwise outranking assessments which are then combined to produce a partial or a complete ranking of alternatives.

The principle, upon which outranking methods are based, is a generalisation of the concept of 'dominance' (mentioned in Section 2.6). An alternative 'outranks' another alternative if, taking into account all available information and decision-makers' preferences, there is strong enough evidence to support the view that the alternative is at least as good as the other alternatives and no evidence to suggest that it is worse (Belton & Stewart 2002). Outranking models are often used when measurement scales for the criteria vary over wide ranges and/or when the units are incomparable (that is, some criteria may be on an ordinal scale and others may be on a categorical or natural cardinal scale).

There are various methods for establishing pairwise outranking assessments and for determining how these assessments will be combined to produce an overall preference ranking. One such method is the ELECTRE method.[55]

With ELECTRE methods preferences are modelled by using binary out-ranking relations. For example, with two criteria *a* and *b*, there are four possible outcomes: *a* is strictly preferred to *b*, *b* is strictly preferred to *a*, *a* is indifferent to *b* or *a* and *b* are incomparable (Figueira et al. 2005). The ELECTRE method uses these outranking relations to define a subset of alternatives. The aim is to obtain as small a subset as possible from which the alternatives can be ranked. The following simple example illustrates how outranking relations are used to

[54] The concept of outranking was devised by Roy (1968).
[55] The family of ELECTRE methods include ELECTRE I, ELECTRE 11, ELECTRE III, ELECTRE IV and ELECTRE Tri. The methods differ in terms of complexity, information required and the problem being addressed (Belton & Stewart 2002).

define a set of alternatives, using concordance and discordance indices (which will be explained shortly).

Table 2.2: Example of the ELECTRE method

	Availability of staff	Ease of set up	Access from the US	Quality of life
Weights	6	6	4	3
Brussels	L	Av	Av	VH
Amsterdam	L	VH	VH	Av
Paris	Av	VL	H	Av

Source: Business location example adapted from Belton & Stewart (2002, p 235)

As can been seen in Table 2.2, there are three alternatives (Brussels, Amsterdam and Paris) and four criteria (the top row). Each criterion is arbitrarily assigned a weight reflecting its relative importance (second row). The alternatives (i.e. the countries) are 'rated' on a five point scale – Very Low (VL), Low (L), Average, (Av), High (H) and Very High (VH) – with respect to the four criteria. Next, a concordance index is calculated between each pair of countries. The concordance index (which lies between 0 and 1) measures how much one alternative is at least as good as another. For example, the concordance index between Brussels and Amsterdam is 0.47. This is calculated by adding together the criteria weights where Brussels is 'at least as good as' Amsterdam and dividing by the sum of *all* the criteria weights. Referring to Table 2.2, it can be seen that Brussels is at least as good as Amsterdam for the criterion 'availability of staff' (6 points) and is better than Amsterdam for the criterion 'quality of life' (3 points). Therefore the concordance index is 6+3/6+6+4+3=0.47. On the other hand, the concordance index between Amsterdam and Brussels is 6+6+4/6+6+4+3=0.68. A higher concordance index indicates that a greater proportion of the criteria for an alternative are at least equal to, or above, another alternative.

The discordance index measures the degree to which one alternative is strictly preferred to another. For example alternative *a* might be at least as good if not better on all the criteria compared with alternative *b*, except for one criterion. However, if for that one criterion alternative *b* performs substantially better than alternative *a* then there is evidence that

alternative *a* might not be better than alternative *b*.[56] When the criteria have been 'rated' on a cardinal scale, the discordance index is the ratio of the maximum weighted value by which one alternative is better than another, and the maximum weighted difference between any two alternatives on any criterion (Belton & Stewart 2002). The discordance index lies between 0 and 1 with scores closer to one indicating that on at least one criterion, *b* performs substantially better than *a*.

When the alternatives are rated using an ordinal scale, as in this example, the discordance index can be calculated by defining a threshold for each criterion, based on a maximum difference between the ratings. For example, in Table 2.2 a five point scale is used to rate the criteria (Very Low, Low, Average, High and Very High). If the threshold is set at say two scale points, it means that Amsterdam cannot outrank Brussels (as indicated by the concordance index) if Brussels is two or more points higher on the scale than Amsterdam for any of the criteria. As Brussels is rated two points higher on the criterion 'quality of life' (Very High/Average), Amsterdam cannot outrank Brussels.

Decision-makers then set concordance and discordance thresholds to establish a subset of alternatives. For example, a concordance threshold could be set at 0.75 and a discordance threshold at 0.25. Alternatives that are outranked by at least one alternative based on these thresholds are not included in the subset.

An advantage of the outranking method is that alternatives that perform badly on one or more criteria are downgraded. This fits with the political reality in the context of health care prioritisation, that technologies which perform badly even on one criterion, are likely to be unacceptable to the public (Dodgson et al. 2009). However, the outranking method relies on decision-makers arbitrarily assigning weights and threshold parameters which can create inconsistencies in decision-making.

2.8 Why PAPRIKA?

The scoring methods explained in the last section each have their own advantages and disadvantages. In this section, the PAPRIKA method is compared to the other methods with

[56] Outranking models are known as 'partially compensatory' as they allow poor performance on some criteria to be compensated for by high performance on others but they do not take into account the magnitude of the difference.

respect to several considerations such as the elicitation method used and how the weights are derived. The main considerations are summarised in Table 2.3.

Table 2.3: Comparison of scoring methods

	SWING/ SMART	DCE/ CA	ACA	AHP	PAPRIKA	Outranking
Elicitation method	Allocate points from least preferred to most preferred; Consider all criteria at same time	Choice based; Usually 4+ criteria in each scenario; two or more scenarios	Choice based; Self-explication; Computer based interview	Pairwise comparisons; Ratio judgements on a nine point scale	Pairwise comparisons/ choice based	Assign weights; Pairwise comparisons based on "at least as good as"
Number of judgements required	Minimum number of judgements	Often limited number of choice sets to reduce overload	Usual time is 45 mins, depending on number of choice sets	Depends on number of attributes	Depends on number of criteria/ levels but likely to be more than DCE/CA/AHP	Depends on number of criteria
Points/ weights derived	Direct Assign weights	Indirect Statistical analysis	Indirect Statistical analysis	Indirect Mathematical algorithm	Indirect Mathematical algorithm	Direct/indirect Assign weights and thresholds
Online individual surveys	No	Yes	No	Yes	Yes	Yes
Individual weights	Yes	No	No	Yes	Yes	No
Validity/ reliability	Arbitrarily assign points	Limited number of choice sets presented	Self-explication; Decision-makers have different choice sets	No attribute levels (e.g. small, medium, large) Consistency ratio	Less decision-maker burden Check for consistency	No levels; Arbitrarily assigned weights and thresholds

Apart from the outranking method, all the scoring methods in Table 2.3, including PAPRIKA, are based on the simple additive model. Baltussen (2006) argues that as MCDA involves trade-offs between criteria, compensatory methods such as the simple additive model (where high scores on one criterion can compensate for low scores on another) are more suitable for

eliciting decision-makers' preferences than non-compensatory methods.[57] According to Hastie & Dawes (2010, p 60 & 67) "the mind is in many essential respects a linear weighting and adding device" and therefore the simple additive model "provide[s] an effective method to predict our own evaluations and preferences". However, an advantage of the outranking method is that alternatives that perform badly on one or more specific criteria can be downgraded.

With the PAPRIKA method decision-makers are asked to choose between two choice sets defined on two criteria[58] whereas outranking, SWING/SMART and some CA methods use ranking, direct rating and/or arbitrarily assigned thresholds to score criteria and rank alternatives. Although the techniques are simple to implement, ranking, direct rating and assigning thresholds relies heavily on the judgement and agreement of 'experts' (or in some cases, members of the public) to appropriately assign scores to the criteria.

Asking decision-makers to rate criteria and/or choice sets can also lead to confusion, inconsistencies and misinterpretation of the data. For example, if a group of people were asked to rate five criteria on a scale of 1 to 10, each individual may interpret the scale differently. One person might consider a '3' to *fairly* low whereas another person may consider it *extremely* low. Someone might simply rank the criteria on an ordinal scale (1st, 2nd, 3rd...). Someone else might assume that the difference between 2 and 4 is the same as the difference between 6 and 8 (an interval scale) or that a score of 8 is twice as important as a score of 4 (a ratio scale) while others may not. How the scores are interpreted depends on the intent and understanding of the decision-makers (Forman & Selly 2001).[59]

Choosing between two alternatives is a cognitively less demanding task than having to rate, rank or score alternatives. "The advantage of choice-based methods is that choosing [ordinal],

---

[57] As mentioned before, outranking models are known as 'partially compensatory' as high performance on some criteria can compensate for poor performance on others but the magnitude of the difference is not taken into account.

[58] Respondents can be presented with choice sets consisting of more than two criteria at a time but as discussed in Section 2.7.4, this is not necessary to produce an overall ranking which is highly correlated with the true ranking.

[59] When considering the use of numbers in scoring and weighting criteria, Stevens' (1946) four levels of measurement – nominal, ordinal, interval and ratio – is useful to consider. Nominal numbers are used to identify an object and convey no meaning about ordering; for example, a telephone number. Ordinal numbers entail a rank or order; for example, 1st, 2nd, 3rd. Ordinal numbers do not imply anything about the strength of difference between the ranks whereas interval numbers do; for example, the interval between 3 and 5 is the same as between 28 and 30. Ratio numbers have the highest level of measure. With ratio scale data, the same ratio between two sets of numbers is considered equal (for example, the ratio between 50 and 25 is the same as the ratio between 4 and 2) (Forman & Selly 2001).

unlike scaling [cardinal], is natural human task at which we all have considerable experience, and furthermore it is observable and verifiable" (Drummond et al. 2005, p 145).

In addition, because decision-makers are not being asked to trade-off the criteria, they may choose to give each criterion the same score or to rank them equally. Choice modelling, on the other hand, forces decision-makers to make trade-offs between criteria.

The AHP method requires decision-makers to make pairwise comparisons of the criteria and/or alternatives at every hierarchical level. Decision-makers are required to make *relative* judgements between two considerations (for example, how many times more important is option A than option B?). Making *relative* judgements is cognitively more difficult than making absolute *ordinal* judgements. Moshkovich et al. (2002, p 276) argue that "eliciting preference information in ordinal form instead of cardinal form will result in more stable and reliable responses".

DCE/CA and ACA, like the PAPRIKA method, elicit preference information in ordinal form. However, typical DCE/CA and ACA surveys present decision-makers with *two or more* choice sets consisting of *more than two* criteria in each choice set. For example, Ryan & Gerard (2003) reviewed 34 DCE studies and found that researchers considered choice sets of between 4-6 attributes to be acceptable in terms of choice complexity. Ranking pairs with just two criteria is less difficult and cognitively demanding than ranking pairs with several criteria (Hansen & Ombler 2008).[60] When there are a large number of criteria to consider at one time, the choice becomes more complex for the decision-maker. Decision-makers might simplify the decision process by focussing on only one or two of the criteria, and as a result, the estimated criteria weights may be inaccurate (Sampietro-Colom 2008, Cameron & De Shazo 2011).

However, the number of choices decision-makers are required to make with the PAPRIKA method is likely to be greater compared to other methods. For example, a value model consisting of six criteria with three levels requires an average of 35[61] choices using the

[60] Charron & Koechlin (2010) monitored the brain activity of 32 volunteers undergoing letter-matching tests and found that the brain was easily able to switch between two hemispheres when carrying out dual functions but when a third activity was added, accuracy was diminished. According to the authors, the results explain why people are better at making choices involving two factors rather than multiple factors.

[61] Given 10% of the explicitly ranked pairs are tied.

PAPRIKA method whereas with traditional methods such as CA, an average of 11 choices is required[62] (Hansen & Ombler 2008).

One reason why a smaller number of choices is required with DCE/CA is that the number of choice sets presented to decision-makers is often very small in relation to the total number of scenarios that could be presented (Bryan & Dolan 2004). Bryan & Dolan found that it is common for DCEs to present decision-makers with eight or nine pairwise comparisons when the total number of possible scenarios might range from 250 to 500 depending on the number of criteria and levels. Presenting decision-makers with a fraction of all available choice sets reduces 'respondent fatigue' but it can also lead to unreliable results. Presenting decision-makers with the same 'sub-set' of scenarios assumes that the decision-makers share similar preferences. Though the number of scenarios presented to decision-makers with an ACA survey is limited because of self-explication, it is problematic to aggregate preferences when different choice sets have been used.

The criteria weights represent the relative importance of the criteria and the willingness of a decision-maker (or a sub-group of decision-makers or the entire sample) to trade one criterion for another (marginal rate of substitution). Apart from the SWING/SMART and outranking methods where decision-makers directly assign points to criteria, AHP and PAPRIKA are the only methods that generate individual criteria weights for every decision-maker. (Typically with methods such as DCE/CA and ACA estimation procedures such as probit, logit and multinomial logit are used to produce a set of weights for the entire sample.) With the PAPRIKA method, the criteria weights of one decision-maker can be compared with another because they have traded-off the same criteria. Overall sample weights are obtained by simply averaging the individual weights across decision-makers. However, with the AHP method, criteria weights can only be directly compared if decision-makers have used the same factors and/or hierarchies (Bolloju 2001). How the weights are aggregated depends on whether decision-makers wish to combine their judgements and at which hierarchy (Saaty 2008). If 'experts' have used their own hierarchies to generate criteria weights AHP combines their outcomes by taking the geometric mean.[63] Further, the 'experts' themselves can be ranked according to their expertise so that their individual evaluations are given more or less importance before calculating the geometric mean (Saaty 2008).

[62] Based on n-1+n(y-2) (Hansen & Ombler 2009).

[63] The geometric mean can be used to calculate an 'average' of two or more groups of values that are measured on different scales. For example, if one criterion is measured on a scale of 0 to 5 and another criterion is measured on a scale of 0 to 10, the geometric mean 'normalises' the ranges so that the individual values can be compared.

Finally, one of the main objectives of this thesis is to discover the preferences of New Zealanders with respect to various criteria for prioritising health care technologies. The most cost-effective and efficient way to survey a random sample of New Zealanders residing throughout the country is to use an online survey. (Online surveys are compared to other methods such as paper-based and telephone interviews in Chapter 5.)

There is a wide range of decision analysis software available. (For a full description of the available software see www.orms-today.org/surveys/das/das.html.) Table 2.4 lists some of the decision analysis software that is supported by PAPRIKA, AHP, CA/ACA and outranking methods.

Table 2.4: Decision analysis software

Software	Supported MCDA method	Pairwise comparison	Time analysis	Sensitivity analysis	Group evaluation	Risk manage-ment	Web-based version
1000Minds	PAPRIKA	Y	N	Y	Y	N	Y
Criterium DecisionPlus	AHP	N	N	Y	N	N	N
D-Sight	MAUT, PROMETHEE	Y	N	Y	Y	Y	Y
DecideIT	MAUT	Y	N	Y	Y	Y	Y
Decision Lens	AHP	Y		Y	Y		Y
Expert Choice	AHP	Y	N	Y	Y	Y	Y
Logical Decisions	AHP, MAUT	Y	N	Y	Y	Y	N
MakeItRational	AHP	Y	N	Y	Y	N	Y
MindDecider	AHP	Y	Y	Y	Y	Y	N
Sawtooth	CA/ACA	Y	N	Y	Y	N	Y
TreeAge Pro		N	N	Y	N	Y	N
Very Good Choice	ELECTRE	Y	N	Y	Y	Y	N

Source: Adapted from Wikipedia and www.orms-today.org/surveys/das/das.html

The only software which supports the PAPRIKA method is 1000Minds (Ombler & Hansen 2012). The software is straightforward to use, the surveys are 'user-friendly' and the software provides data in an easy-to-use format.[64] I was also granted a free licence to use 1000Minds which enabled me to conduct several surveys without having to consider the cost. (The cost of decision analysis software can range anywhere from free to far in excess of US$10,000 (Buckshaw 2010)).

[64] A description of 1000Minds software and how it works is given in Chapter 5.

In summary, I chose to use the PAPRIKA method because it is choice based (i.e. cognitively less demanding than some other methods), it minimises complexity (pairwise comparisons defined on two criteria), and unlike other conjoint methods where alternatives are ranked based on only a fraction of all possible pairwise comparisons, PAPRIKA can be implemented as a full or an incomplete ordinal information system (with minimal difference between the two). The PAPRIKA method also generates individual weights for every decision-maker which are easily aggregated, a feature not available with most other methods. In addition, designing and implementing a decision survey using 1000Minds software is straightforward and cost-effective, and in terms of the survey format, most respondents find it easy to follow.[65]

2.9 Conclusion

Allocating scarce public resources across competing health services is a complicated and difficult process. Numerous factors need to be considered as the decisions made will affect many people both directly and indirectly. There are many different approaches to prioritising health care, some of which have been discussed in this chapter. One of the approaches, MCDA, assists decision-makers in making complex decisions involving multiple factors in an explicit, consistent and transparent way. Within MCDA, there is a variety of preference elicitation and scoring methods, all of which have their own advantages and disadvantages.

In this thesis a MCDA framework is used, in conjunction with 1000Minds software (Ombler & Hansen 2012) and the PAPRIKA scoring method (Hansen & Ombler 2008), to determine the willingness of New Zealanders to make trade-offs in health care.

[65] The survey design is discussed in Section 5.7.

~ Chapter 3 ~

Focus groups

3.1 Introduction

"If there was plenty of money there wouldn't be any worries..."

Member of the retirees' focus group

"But you can't save everyone. You have got to make the decision."

Member of the GP practice focus group

As discussed in the previous chapter, establishing the appropriate decision criteria and associated levels is a very important stage in the MCDA process. The criteria need to include all relevant information, be independent and easy to understand. Involving key stakeholders in the development of the criteria ensures transparency and promotes trust in the process as well as ensuring that the right criteria are captured.

In this chapter and the next, the process of establishing the relevant criteria and levels is explained. This chapter begins with a discussion of the qualitative methods used to establish criteria in MCDA including citizens' juries and focus groups. The use of vignettes in health research and their purpose in this thesis is then explained. The second part of the chapter describes how focus groups are used to incorporate the views of key stakeholders in developing the criteria, concluding with a summary of the focus group meetings. The criteria and levels are presented in Chapter 4.

3.2 Qualitative methods for eliciting criteria

Citizens of a country view their public health system from a variety of perspectives: as voters, taxpayers, patients, health professionals, educators etc. They are interested in how the public health system affects not only themselves but also "their families, neighbours and fellow citizens, both now and in the future" (Lenaghan 1999, p 48). Because health care is publicly funded, it is important to elicit the community's preferences and values with respect to health care prioritisation (Gafni & Birch et al. 1993).

Two approaches can be used to elicit public views: an 'uninformed and undeliberated' approach such as ranking lists of treatments and services, one-off surveys and public consultation, or a 'deliberate' approach such as citizens' juries and group discussions (Dolan et al. 1999). With the 'uninformed and undeliberated' approach, respondents are often not given much time or information and therefore the results obtained may be of questionable validity.[66] For instance, ranking lists of treatments is a fairly rudimentary yet cognitively complex activity; surveys often do not allow enough time for the public to understand and consider the trade-off decisions being made; and public consultation such as public meetings are often poorly attended or can be dominated by interest groups (Lenaghan 1999).

'Deliberate approaches' such as citizens' juries and group discussions are increasingly being used to elicit the general public's preferences (Jordan 1998, Dolan et al. 1999). In the next two sections these qualitative methods are explained and the reasons for using focus groups in this thesis to elicit potential criteria for the decision survey[67] are discussed.

3.2.1 Citizens' juries

A citizens' jury consists of a small group of participants (usually 12-24 people) who have been chosen with the intention of being broadly representative of the overall population. They meet for a specified period, usually around three days, during which time they are presented with 'evidence' by experts. Their brief varies: they may be asked to debate policy or to decide on funding priorities for example. Jurors have the opportunity to ask questions and debate the issues. They are exposed to a wide range of viewpoints and are given time and information to fully deliberate before making a collective decision.

Lenaghan (1999) reviewed four citizens' juries ranging in scope from deciding which of four palliative care treatments to fund to setting broad health priorities in a publicly-funded health system. Lenaghan found that the jurors were able to manage specific tasks such as choosing between two health programmes much better than deciding on broader tasks such as financing the health system. The majority of participants agreed that the wider general public should be involved in rationing decisions.

---

[66] A counterview in the marketing literature is that 'top of mind' questioning reveals respondents' true feelings as opposed to deliberated responses where respondents might try to give the 'correct' answer.

[67] An online decision survey [implemented through 1000Minds software (Ombler & Hansen 2012)] is used to estimate the preferences of the general public with respect to the six criteria. The survey is explained in Chapter 5.

Although a citizens' jury is a useful tool for involving the public in debates about the values or criteria that should be used when prioritising health care (Lenaghan 1999), a citizens' jury can be expensive to set up. It is also arguable whether such a small group (12-24 people) is able to adequately 'represent' the wider population (i.e. whether the results from the citizen's jury can reasonably be extrapolated to the wider population.)

3.2.2 Focus groups

A focus group consists of a small group of people who have been gathered together for a group discussion in order to gain insight into a particular topic (Kreuger 2000). The group's discussions are facilitated by an interviewer or moderator who guides the group in a structured or unstructured way, depending on the purpose of the focus group (Denzin & Lincoln 1994).

A crucial objective in focus group research is discovering *why* participants hold certain opinions or beliefs. This is achieved through interaction of the group participants – participants can question each other, seek clarification, and discuss and share their thoughts thereby providing insights into the different perspectives of participants within and between groups (Brondani et al. 2008, Casey & Krueger 2004). Other qualitative methods such as individual interviews tend to focus on *what* people think rather than *why* they think it.

The 'safety in numbers' aspect of focus groups encourages participants to express their ideas and to seek clarification (Kitzinger 1994). On the other hand, poor group dynamics or one or two dominant group members can result in participants being reluctant to volunteer information or the group focussing on specific viewpoints only. When this happens a false sense of consensus can arise if participants are reluctant to contradict other members who have a conflicting opinion especially when the topic is unfamiliar (Brondani et al. 2008).

Another issue with focus groups is representativeness. Because the number of participants in focus groups is typically small and participants are often recruited by convenience sampling[68] the results from focus groups may not be generalisable. However, with good facilitation, focus groups can provide valuable information at a relatively low cost.

[68] With convenience sampling, participants are selected based on their knowledge of the subject or ease of access to the researcher. The size of focus groups and the different methods of sampling are discussed in Section 3.4.

The primary purpose for using a qualitative method in this thesis is to elicit potential criteria for the decision survey. Focus groups are used rather than a citizens' jury for several key reasons. Conducting several focus groups instead of one citizens' jury enables more people to be involved. As focus group meetings typically run for one to two hours, it is easier to recruit participants for shorter meetings than for a citizens' jury which can run for two to three days. If focus groups participants are already associated with each other in some way, they may feel more comfortable in offering their opinions.[69] In addition, discussions can be facilitated by the use of health vignettes (which will be explained later) rather than hearing 'evidence' from experts. Although a citizen's jury is more structured and has the advantage of hearing 'evidence' from experts, for the purposes of this thesis, focus groups are more suitable.

Focus groups have been used extensively to elicit the views of the general public with respect to service provision and the health care needs of communities (Barbour 1999). A summary of several health research studies using focus groups is presented below.

3.2.3 Literature review of research involving focus groups[70]

Using focus groups as part of a mixed-methods approach in social and health sciences research, has become increasingly popular. Not only is the information obtained from focus groups insightful on its own, but it can be used as a basis for the quantitative component of a study (Kidd 2000, Barbour 2005, Rabiee 2007).

Dolan et al. (1999) explored whether people's views on priority-setting in health care change, when they are given time to discuss and deliberate. Ten groups, with 5-7 people in each group, met twice. In the first meeting participants filled in a questionnaire about setting priorities in health care. After a discussion, they ranked hypothetical groups of patients based on quality and/or length of life. In the second meeting the participants ranked four patients and as a group made three pairwise comparisons between several groups of patients. Dolan et al. (1999) found that the participants' views changed between the meetings. For example, half of the participants initially gave lower priority to smokers, heavy drinkers and illegal drug users but after discussion, many respondents changed their minds and no longer assigned

[69] In this thesis, focus groups were established based on a commonality such as age, workplace or occupation for instance. Participants on citizens' juries are chosen to broadly represent the target population and therefore are unlikely to know each other or have a strong association.

[70] In this section several studies that have used focus groups to explore priority-setting in health care are reviewed. The objective is to illustrate how focus groups are a good way to elicit potential criteria for priority-setting. The actual criteria elicited in these studies will discussed alongside the potential criteria from the focus groups in Chapter 4.

lower priority to these groups. The authors' main conclusion was that if the 'considered opinions' of the general public are required then participants need time to discuss and ponder the issues, and that surveys that do not allow this to happen may be of questionable value.[71]

The Somerset Health Authority in the UK used focus groups to consult with the public about health service priorities. They established eight focus groups with 12 people in each group who met three times a year to discuss issues concerning the health authority. Bowie et al. (1995) reviewed the Somerset Health Authority's approach to public consultation. They found that this approach was successful for developing a consensus on broad values which were "representative, valid and focused on community rather than individual values" (p 1155).

Focus groups were used in Sampietro-Colom et al.'s (2008) study to ascertain and define criteria for prioritising patients in need of joint replacements. Four groups were established, with 5-10 participants in each group, representing the general population, patients and close relatives, allied-health professionals, and consultants. Each group's discussion was used to identify potential criteria. Participants then ranked these criteria from least important to most important. Of the 15 criteria identified, seven were chosen as relevant prioritisation criteria. Each criterion was demarcated into levels and incorporated into different patient scenarios for use in a conjoint analysis survey.

Jan et al. (1999) conducted six focus groups to determine what attributes should be included in a conjoint analysis survey about public hospital services in South Australia. The focus groups were held in a mix of metropolitan and country areas, with participants chosen from a wide range of socio-economic groups. To generate discussion and encourage participants to think about their health services, they were asked if they had any problems with their local health services. They were also asked what factors would encourage them to use one hospital service over another. Although many factors were suggested, they could be grouped into four main 'attributes'. These 'attributes' were included in a mailed questionnaire completed by 231 respondents.

Stronks et al. (1997) established separate panels comprising patients, the general public, general practitioners, specialists, and health insurers with approximately nine people in each

[71] By analysing the groups' discussions and the groups' rankings of the hypothetical patients, the authors also discovered which of the ethical principles of health care rationing (identified from the literature) the focus groups supported most strongly. These were: 'rule of rescue', health maximisation and equalisation of lifetime health.

group. The panels were presented with 10 descriptions of health treatments and asked which should be funded if only one third of the total budget was available. The authors found that there were substantial differences in the way the different groups approached the issue of which services should be funded. The main aim of the study was to gain insights into the decision-making process rather than to determine which treatments to fund.

The research surveyed in this section demonstrates that information gleaned from focus groups is valuable: participants have the time and opportunity to discuss their views, opinions are sought from representative groups rather than individuals, and when focus groups are used as part of a mixed-methods approach they can add validity and strength to the quantitative part of the research.

As mentioned earlier, the main purpose of the focus groups is to elicit potential criteria for the decision survey. Vignettes are used in the focus groups as a vehicle for introducing the topic of health care prioritisation and to stimulate discussion. The use of vignettes in health research and the 14 vignettes used in this thesis are explained in the next section.

3.3 Vignettes in health research

According to Spalding & Phillips (2007), vignettes have been documented as being useful in research for more than 25 years. A vignette is usually written in the format of a short story or case study where hypothetical characters are described. Vignettes are a form of 'indirect' or 'third person' questioning as participants are asked questions based on a hypothetical scenario rather than being asked direct questions. Because the vignettes are based on hypothetical situations (rather than *actual* ones), participants feel more at ease in discussing their feelings and attitudes. Vignettes are therefore useful in exploring potentially sensitive topics in a non-threatening way. Hughes & Huby (2002) concluded that, "Vignettes, used alone or in conjunction with other research techniques, can be valuable research tools in the study of people's lives, their attitudes, perceptions and beliefs" (p 385).

In Berney et al.'s (2005) study, a multi-stage method was used to discover how GPs applied ethical principles when allocating scarce resources. The GPs involved in the study were interviewed and asked to identify key resource allocation issues. The main ethical issues were written up in the form of vignettes describing hypothetical patient case studies. These vignettes were then discussed in a series of focus group meetings with GPs. The use of

hypothetical case studies enabled the GPs to discuss and debate sensitive and controversial issues in a non-threatening environment. The authors found that this approach developed greater levels of trust between the facilitator and the GPs compared with simple one-off interviews, provided more in-depth and open discussion, and allowed examination of areas of agreement and disagreement concerning ethical principles.

Brondani et al. (2008) found that using vignettes in focus groups encouraged participants to talk about their personal experiences, which promoted good group interaction with very little disagreement. In Brondani et al.'s (2008) study six focus groups were formed with up to nine older men and women in each group. Two short situational vignettes portraying realistic scenarios were used to promote group discussion. Focussing on the vignettes allowed participants to discuss comfortably what the authors considered to be a sensitive subject, oral health.

Schoenberg & Ravdal's (2000) study explored the awareness and attitudes of older adults towards the use of formal community-based or home-health services aimed at keeping elderly people at home rather than in institutions. The authors felt that assessment methods such as pre-structured questionnaires often contained investigator bias and contributed to the paucity of information in this area. They developed, pre-tested and administered narrative-style vignettes with follow-up questions to 115 elderly people to explore their attitudes towards community based services. They concluded that the vignette approach for data collection and analysis was an appropriate tool for awareness and attitudinal research.

Vignettes are used in this thesis to help facilitate discussion and debate in the focus groups and to provide consistency and comparability between the groups.

3.3.1 Vignettes in this thesis

The primary purpose of using vignettes in this thesis is to encourage discussion in the focus groups with the intention of eliciting participants' thoughts on prioritising health care. Before attending a focus group meeting participants receive a link to an online ranking survey where they are asked to rank 14 health vignettes in order of their perceived benefit or value to society.[72] According to Barbour (1999), although ranking exercises can be considered an

[72] 1000Minds software (Ombler & Hansen 2012) was used for the ranking survey. The survey is explained shortly.

artificial exercise, they are useful for eliciting the factors that influence decisions or attitudes towards to specific issues.

After talking with several health professionals and ascertaining what health treatments were of current interest (e.g. the cervical cancer vaccine) I established 14 vignettes which encompassed a variety of health treatments ranging from preventive treatments (e.g. statins) through to life saving treatments (i.e. dialysis). The reason for choosing a diversity of vignettes was to encourage discussion within the focus groups thereby enabling as many potential criteria to be elicited as possible. The 14 vignettes are presented in Figure 3.1.

Information for the health vignettes was gathered through reading secondary research, current reports and studies on the treatments, and from discussions with health professionals. The vignettes are described at the treatment level rather than at an individual patient level to encourage a societal perspective. Care was taken to ensure that the vignettes were as concise and as consistent as possible. Each vignette follows a similar format with respect to lay-out, language used, information included and length of description. This consistency ensured that participants easily understood the vignettes and were not swayed by emotive language, inconsistent information or difficulty in interpretation. The vignettes were checked for accuracy and clarity by an expert in each area.[73]

Several main criteria are commonly associated with prioritising health treatments (e.g. increase in quality and/or length of life). These criteria form the basis of the information contained in the vignettes. (A literature review of the general principles as well as the specific criteria that could be included in a health prioritisation process is presented in Chapter 4.) Each vignette includes: a brief description of the treatment, the reasons for treatment, the effectiveness of the treatment, the number of people to be treated, possible side effects and, where relevant, the age, gender and ethnicity of a representative patient.

The cost of a treatment is not included in the vignettes for several reasons. First, cost is an obvious criterion – the purpose of a prioritisation framework is to consider the *costs* and the benefits of treatment (alongside any other additional considerations) when allocating resources across a range of health services. The benefits of treatment and other factors, however, may not be so obvious.

[73] The health professionals who had input into developing the vignettes are mentioned (and thanked) in the Acknowledgements.

Second, when cost is included in the vignettes, participants need to consider how many patients are being treated and not just the total cost, as each vignette differs in terms of how many patients are being treated. Estimating the cost effectiveness of the different treatments could be difficult for some participants. For example, comparing 7000 hip replacements costing $119,000,000, with dialysis for renal disease (440 patients) costing $22,000,000 with abatacept for rheumatoid arthritis (30 patients) costing $900,000.

Third, cost will not be included as a criterion in the decision survey but will be considered alongside the benefits of treatments within the prioritisation framework (in Chapter 9). Cost will not be included in the decision survey as respondents are asked to choose between two hypothetical patients instead of two health treatments. This means that cost needs to be 'cost per patient' but this is not always practicable (e.g. vaccines that are provided as an overall programme). Also, including cost as a criterion adds complexity to the trade-off questions and creates uncertainty relating to opportunity cost (Bryan et al. 2002). These points are further discussed in Chapter 4.

Figure 3.1: Health vignettes

Antiretroviral drugs for HIV

- Human immunodeficiency virus (HIV) is a virus that affects the immune system, and can lead to AIDS within 8-10 years. AIDS-related illnesses include eye infections, pneumonia, thrush, skin cancer and brain tumours.
- HIV is transmitted through sex, blood transfusions, sharing of needles and between a mother and baby during pregnancy, birth and breastfeeding.
- Men and women can be heterosexually infected with HIV, although men who have sex with men are most at risk.
- A combination of at least 3 antiretroviral drugs can be used to suppress the HIV virus and control its progression.
- There is no cure for HIV but antiretroviral treatment can increase life expectancy by an average of 13 years.
- Of the 180 people diagnosed with HIV each year in NZ, approx 125 will start antiretroviral treatment.
- Number of people to receive antiretroviral drugs: 125 for the rest of their lives

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Vaccine for preventing cervical cancer (Gardasil)

- Cervical cancer is caused by the human papillomavirus (HPV), a common virus passed on by sexual contact.
- Gardasil is a vaccine that targets HPV types 16 and 18 which cause up to 70% of cervical cancer and HPV types 6 and 11 which cause 90% of genital warts.
- Gardasil is given to females 12-18 years and is most effective when girls have not been sexually active.
- Although cervical smears will still be needed there will be a reduction in diagnosis and treatment costs for abnormal smears and for genital warts.
- Each year about 160 women are diagnosed with cervical cancer and 60 will die.
- In the future Gardasil will prevent around 30 deaths each year.
- Number of young women to be vaccinated: 50,000 over 5 years

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Dialysis for End-Stage Renal Disease

- End-stage renal disease is when the kidneys no longer function well to enough to keep a person alive and renal replacement therapy (RRT) is required.
- RRT includes kidney transplantation, haemodialysis and peritoneal dialysis.
- Dialysis removes waste and extra fluids from the blood using a special filter (haemodialysis) or a catheter in the abdomen (peritoneal dialysis).
- Dialysis is time-consuming and is done in hospital or at home.
- The major causes of renal failure are diabetes, kidney disease, high blood pressure and genetics.
- The average age of a dialysis patient is 56 years, with many patients over 65. Almost 50% of patients are Maori.
- The number of people receiving dialysis could double in the next 5 years.
- Approx 50% of people starting dialysis are still alive after 5 years.
- Number of people to start dialysis: 440 for the rest of their lives

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Growth hormone treatment for Prader-Willi Syndrome

- Prader-Willi Syndrome is a rare genetic disorder, which causes low muscle tone, developmental delay, behavioural problems, and an insatiable appetite and obsession with food which leads to life-threatening obesity.
- Growth hormone treatment (GHT) builds bone density and muscle tone, increasing height and boosting energy.
- Children gain the most benefit when given GHT while they are still growing.
- If left untreated, children will end up 12-36 cm shorter than the average adult height.
- There is no known cure, although GHT can improve children's short-term growth and/or their final height.
- Prader-Willi Syndrome occurs in approx 1 in 25,000 births. Currently 23 children in NZ have the syndrome.
- Number of children to receive growth hormone treatment: 3 until they reach adulthood

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Hip replacements

- A hip replacement is a surgical procedure in which the damaged hip joint is replaced by a prosthetic implant.
- Hip damage is caused by osteoarthritis, rheumatoid arthritis and hip fractures.
- The most common cause of deterioration of the hip joint is osteoarthritis. As the cartilage lining becomes damaged and wears away, the bones within the joint rub together causing pain and making it difficult to get around.
- It can affect men and women, and is more common over the age of 50.
- A hip replacement relieves pain and restores function to the joint. Patients become mobile again and can lead a normal lifestyle.
- A hip replacement typically lasts 15-20 years.
- Number of people to receive a hip replacement: 7000

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Imatinib mesylate for chronic myeloid leukaemia

- Chronic myeloid leukaemia (CML) is a rare blood cancer. People with CML are more prone to infections and have an increased risk of bleeding.
- The cause of CML is unknown.
- Imatinib mesylate is a drug that blocks or switches off a protein which instructs the body to keep producing abnormal blood cells.
- CML affects women and men, most commonly between the ages of 45-55 years.
- People diagnosed with CML usually live for around 5 years.
- Successful treatment with imatinib mesylate can increase life expectancy from 5 years to over 10 years.
- There are 1-2 new cases per 100,000 people each year.
- Number of people to receive imatinib mesylate: 40 for the rest of their lives (approx 10 years)

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

IVF Treatment

- Infertility is when a couple is unable to get pregnant after a year of trying. It may be unexplained or caused by factors such as endometriosis, blocked fallopian tubes or poor quality sperm.
- The grief experienced as a result of childlessness is similar to clinical depression.
- 25% of couples experience infertility within their reproductive life time, affecting men and women almost equally.
- The most successful infertility treatment is In Vitro Fertilisation (IVF). IVF is when eggs are fertilised outside of the body and then re-implanted into the mother.
- Of the 2450 women in NZ who receive infertility treatment each year, 50% choose IVF treatment.
- 80% of women (under 37 years) completing IVF treatment will have a baby.
- Number of women to receive complete IVF treatment: 1225

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Methadone for opioid addiction

- Methadone is used to treat people who have an opioid addiction (e.g. heroin or morphine), by helping them to reduce their use of opioids.
- Methadone reduces the death rate from overdoses and the spread of infectious diseases (hepatitis B, C or HIV from injecting drugs) and improves the health of addicts.
- Opioid addiction is also associated with high cannabis and tobacco use, low health status and low rates of employment.
- Methadone treatment reduces the substantial social and economic costs resulting from drug abuse.
- Alternatives to methadone such as abstinence based treatments are largely ineffective.
- Relapsing is common with methadone treatment. 98% of addicts stop injecting drugs after an average of 5 years' stabilisation.
- Number of people to receive methadone: 4000 (until they stop their opioid use).

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Positron Emission Tomography (PET Scan)

- A PET scan is a sensitive form of x-ray scanning which uses small amounts of radioactive material to detect diseases such as cancer, some heart disease and brain abnormalities.
- PET scans are most commonly used to detect cancer to determine if it has spread, and to assess the effectiveness of treatment. They help clinicians plan the best form of treatment, e.g. surgery, chemotherapy or palliative care.
- PET scans can save people's lives by providing a more accurate diagnosis.
- The costs and trauma of major surgery for patients who cannot be cured can be avoided.
- About 16,000 people in NZ develop cancer each year.
- 75% of patients who receive a PET scan have their treatment changed as a result.
- Number of people to receive a PET scan: 5000

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Oral drugs for erectile dysfunction (e.g. Viagra, Cialis)

- Erectile dysfunction (ED) occurs when a man is unable to maintain an erection.
- Most ED cases are caused by physical problems, with a small percentage caused by psychological problems.
- Physical problems include high blood pressure, high cholesterol, diabetes, stress, smoking and excessive alcohol intake.
- ED can lead to loss of confidence and self-esteem or depression, as well as to relationship problems.
- Oral drugs such Viagra or Cialis usually allow an erection to occur (with stimulation).
- Around 40% of men over 40 will have erection problems but only 5% will seek help.
- The success rate is approx 60%.
- Number of men to receive 1 year supply of drugs: 9000 (approx 2.5% of men with ED)

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Statins for patients at high risk of cardiovascular disease

- Cardiovascular disease (heart, stroke and blood vessel disease) is the leading cause of death and hospitalisation in NZ.
- Risk factors are smoking, physical inactivity, an unhealthy diet, high cholesterol, high blood pressure and diabetes.
- Death rates are higher for men than women and are much higher for Maori and Pacific Island people.
- Statins are drugs that reduce the production of cholesterol by the liver, helping to prevent blood vessels becoming blocked with fatty deposits.
- Approx 20% of people over the age of 35 could benefit from using statins, depending on the threshold for absolute risk.
- Statins reduce the risk of a heart attack or coronary death by about a third.
- Number of people to receive statins: 220,000 (for the rest of their lives, potentially preventing 66,000 heart attacks or coronary deaths)

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Abatacept for last-line treatment of rheumatoid arthritis

- Rheumatoid arthritis (RA) is a chronic and progressive disabling disease that causes pain and joint inflammation and can cause joint damage.
- Onset of RA mainly occurs between 40-70 years, affecting 3 times as many women as men.
- Abatacept helps stop the immune system attacking healthy tissues in the body.
- Abatacept is not a cure for RA but when combined with other drugs can significantly improve the quality of life of a person by reducing pain, joint inflammation and damage to bones and cartilage.
- Abatacept is used when treatment with other drugs has been unsuccessful.
- A serious side effect is that it can reduce a person's ability to fight infection.
- Number of people to receive abatacept: 30 for the rest of their lives

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Hand sanitiser use in primary schools

- Hand washing helps reduce infectious disease transmission. An alcohol-based no-rinse hand sanitiser is an alternative to using soap, water and drying facilities.
- It helps to reduce the spread of respiratory and gastrointestinal infections by killing various types of bacteria and inactivating different kinds of viruses.
- On average, approximately 11% of children are absent from school each week due to illness.
- In addition to children being ill, spread of the illness harms other pupils, staff and caregivers. Also parents/caregivers may require time off work due to illness or caring for a sick child.
- Alcohol-based hand sanitisers in schools could reduce the rate of absenteeism due to illness by 20%-50%.
- Number of children to use hand sanitisers: 400,000 (for one 4 month period during winter)

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

Service for postnatal depression

- Postnatal depression (PND) is when mothers experience feelings of anxiety, irritability and hopelessness that do not improve. It can occur at any time during the first year after giving birth.
- PND can result in longer-term cognitive, emotional and developmental problems in the baby because the mother is less likely to bond with the baby and provide a safe, nurturing environment.
- PND affects about 13% of new mothers and causes stress for partners, friends and family.
- If untreated, PND can go on for several months or years and can lead to severe depression.
- Treatment options include additional support and social contact, medication, natural remedies, counselling and psychological help.
- Currently only 3% of the most severe cases are accessing mental health services.
- Number of women to receive treatment: 2500 for up to one year (approx 30% of women with PND)

When ranking this treatment don't consider its cost - just consider its benefits/value to society.

To ensure that the vignettes were easily understood by potential focus group participants and suitable for the task of encouraging discussion in the focus groups a pilot test was conducted. A sample of 27 participants, selected by convenience sampling, was asked to rank the 14 vignettes in terms of their overall benefit or value to society.[74]

A screen shot of the online ranking survey is displayed in Figure 3.2.[75] When participants clicked the link to the survey, they were presented with 14 health treatment descriptions (vignettes) and asked to order them in order of importance. The initial order of the vignettes differed randomly across participants to prevent order bias.[76] Participants could read a description of the treatment – just like in Figure 3.1 – by simply clicking on the treatment name. In addition, participants received a PDF of the treatment descriptions (that they could print off if they wanted to) with the invitational email.

[74] Participants were sent an email with a link to the 1000Minds ranking survey. A copy of the email sent to the pilot participants is in Appendix 3.1.
[75] A screen shot is an image taken of the computer screen used for illustration purposes.
[76] Order bias occurs when participants favour particular objects, in this case vignettes, because of their initial placement in a list.

Figure 3.2: Screen shot of the online ranking survey for the pilot study participants

1000Minds software (Ombler & Hansen 2012) produces ranks of the vignettes for every individual as well as the median and the mean ranks of the vignettes for the entire sample. The median and the mean ranks of the vignettes for the pilot study are presented in Table 3.1.

Table 3.1: Median and mean ranks of the vignettes

Vignette	Median rank	Mean rank
Statins for cardiovascular disease	1	3.0
Vaccine for cervical cancer	3	4.4
PET scan	4	5.4
Dialysis for renal disease	6	6.2
Hip replacements	6	6.5
Imatinib mesylate for chronic myeloid leukaemia	7	7.3
IVF treatment	7	8.0
Service for postnatal depression	8	7.1
Antiretroviral drugs for HIV	8	7.3
Abatacept for rheumatoid arthritis	9	8.5
Methadone	10	9.3
Hand sanitiser for use in primary schools	11	9.6
Oral drugs for erectile dysfunction	12	11.2
Growth hormone treatment	13	11.2

In addition, the frequencies of ranks for the vignettes are provided. For example, as can be seen in Figure 3.3, 15 of the 27 pilot study participants ranked ‘statins for cardiovascular disease’ first and 10 of the 27 participants ranked ‘oral drugs for erectile dysfunction’ last. In another example, one person ranked ‘IVF treatment’ first whereas another person ranked it last. For the purpose of the focus groups, the rankings do not matter. What is of interest is finding out *why* one person ranked it first and *why* one person ranked it last. Discussing results such as this in a focus group is good way to initiate discussion and tease out the factors contributing to the rankings.

Figure 3.3: Screen shot of the frequencies of vignette ranks from the pilot study

Frequencies of ranks for each of the 14 technology descriptions

Number of participants (out of 27 in total) who gave the identified technology the identified rank.

RANK	Statins for cardiovascular disease	Vaccine for cervical cancer	PET Scan	Dialysis for renal disease	Hip replacements	Service for postnatal depression	Antiretroviral drugs for HIV	Imatinib mesylate for chronic myeloid leukaemia	IVF Treatment	Abatacept for rheumatoid arthritis	Methadone	Hand sanitiser use in primary schools	Growth hormone treatment	Oral drugs for erectile dysfunction
1st	15		4	2			2		1		1	2	1	
2nd	3	7	5	1	2	4		1		1	2	2		
3rd	1	9	1	4	4	2	3	2				1		
4th	2	2	4		4	1	3	5		2	1			2
5th	1	3	3	5	1	2	1	3	5				1	1
6th	1		1	3	3	3	2	1	5	3	2	1	1	2
7th	1	2	2	2	6	1	2	2	3	3	2		1	
8th	1		2	4		3	2	1	1	4	3	1	2	2
9th		2		3	2	4	3	2	4	3	2	2	1	
10th	1	1		1		3	3	5	2	5	1	4	1	
11th			2		3	1	5	3	1	2	3	2	2	2
12th	1	1		1		1		2	3	3	2	5	2	5
13th			1	1	1	2			1	1	6	2	10	3
14th			2		1		1		1		2	5	5	10

In addition to completing the ranking exercise, the participants in the pilot study were asked whether the vignette descriptions were understandable, whether the survey instructions were clear, and whether the information contained in the vignettes was sufficient to prioritise the treatments. As a result, several minor changes were made. For example, the following instruction was added to the email sent to participants "... drag the green diamonds to alter the order of the health treatments and click on finish when you have completed your rankings." (The instruction was already on the survey page but not all respondents read it.)

The process of establishing the focus groups is explained in the next section.

3.4 Establishment of focus groups

According to Casey & Krueger (2004), having a clear purpose for a focus group study ensures that the planning, conduct and analysis of the focus groups will be much easier. The purpose of using focus groups in this thesis is to discover what factors a cross-section of the general public considered important when prioritising publicly-funded health treatments. These factors will be incorporated into the criteria used in the decision survey.

When establishing and conducting focus groups these components need to be considered: sampling methodology, number of groups, number of participants in each group, facilitating the groups, and analysing and reporting the information from the focus groups. These components are discussed separately in the following sub-sections.

3.4.1 Sampling[77]

An important consideration when using focus groups is that the members are reasonably representative of the population being studied. Participants can be selected to form relatively heterogeneous groups or homogeneous groups or a mixture of the two (Barbour 2005). In marketing research, demographically heterogeneous groups are mainly used, where participants who do not know each other are recruited from a variety of backgrounds to meet and discuss new products or advertising campaigns for example. In contrast, focus groups in health research tend to be demographically homogeneous, with participants sharing similar

[77] Ethical approval (Category B) was obtained from the University of Otago to conduct the focus groups and administer the decision survey. In addition, a mandated response was sought from the Ngai Tahu Research Consultation Committee. The Committee considered the research to be "of importance to Maori health".

characteristics such as gender, age or ethnicity (Krueger 1994). Because focus groups are often confronted with difficult or complex topics, participants who feel comfortable with each other are more willing to discuss, debate and challenge one another, leading to less inhibited discussion and providing a wide range of responses (Barbour 2005, Kitzinger 1994).

Rabiee (2007) and Kitzinger (1995) agree that the concept of homogeneity exploits participants' shared experiences, but that it can be advantageous to assemble a diverse group of participants to encourage honesty and spontaneity allowing for a wider range of perspectives to be explored.

A problem with homogeneous groups that can arise is the impact of pre-existing cliques within the focus group. For example, if a focus group consists of individuals from different levels in an organisation, some members may not feel comfortable sharing their views or challenging opinions (Kitzinger 1995, Barbour 2005).

In his discussion of whether to recruit on the basis of homogeneity or heterogeneity, Barbour (2005) recommends that there needs to be some diversity within a group to stimulate discussion. Regardless of how participants are selected, groups are rarely selected randomly and a group will never be truly homogeneous (Bender & Ewbank 1994, Kitzinger, 1994). For instance, participants recruited for a focus group based on specific characteristics such as age, gender or work place will share certain characteristics but may differ in other aspects.

Participants can be selected for focus groups by random sampling, convenience sampling or purposive sampling. With random sampling, every potential participant has an equal chance of being selected; for example, participants might be selected randomly from the telephone book or the electoral roll. Groups formed by random sampling are more representative of the general population compared with other sampling methods. Potential disadvantages of random sampling are that it tends to be more expensive compared to other sampling methods and the diversity of participants can result in weaker group dynamics.

Convenience sampling is the most common method for selecting participants for focus groups (Stewart et al. 2007). Participants are selected on the basis of ease of access to the researcher (i.e. convenience) and for their knowledge of the subject matter. Care should be taken to ensure, as much as possible, that the groups consist of representative members of the larger

population of interest. Groups formed by convenience sampling are useful for exploratory purposes, such as obtaining a variety of views relating to a specific problem.

With purposive (or purposeful) sampling participants are selected according to specific criteria, usually for in-depth studies on a particular subject (Russell & Gregory 2003). Because the groups are formed to focus on a specific topic, they may not be representative (Rabiee 2007).

Nair et al. (2002) used a combination of purposive and convenience sampling in a study about what patients know about their medications. The views of patients, physicians and pharmacists were sought. Purposive sampling was used to stratify potential participants into the appropriate groups and convenience sampling was used to recruit the participants from naturally occurring groups (e.g. patients).

In this thesis, a similar approach to Nair et al. (2002) was taken to form the focus groups. Purposive sampling was used to select people affected by, and/or interested in, health care prioritisation, and convenience sampling was used to recruit participants for the groups. For example, I contacted a GP and asked whether her practice would be interested in being involved. Subsequently, two GPs, one GP registrar, one practice nurse and the practice manager formed one of the focus groups.

Six focus groups were formed: a GP practice (comprised of two GPS, one GP registrar, one practice nurse and a practice manager), a nurses' group (a mix of public and private sector nurses), a non-medical health workers' group (workers from a non-medical health organisation), a public health group (a mix of workers from Public South Health, the University of Otago Preventive and Social Medicine Department and Plunket), a Maori health provider group (13 workers in various roles within the organisation) and a retirees' group (seven retired people over the age of 65).[78]

Conducting six focus groups enabled a wide range of perspectives to be elicited. Results were compared between the groups with all groups producing similar results.

The theoretically optimal number of participants for a focus group is discussed below.

[78] As discussed later, because participants shared a similar socio-demographic characteristic such as age or occupation for example, they felt comfortable together, even though some participants did not know other members of their group. Discussion flowed, opinions were challenged and valuable insights into health prioritisation were gained.

3.4.2 Size

The appropriate number of participants in a focus group needs to be carefully considered. There needs to be enough participants to be able to generate diverse ideas and create interaction but, on the other hand, too many participants can prevent some respondents from sharing their thoughts (Casey & Krueger 2004).

A review of the literature suggests a recommended group size of between four and 10 participants. Kitzinger (1995) suggests between four and eight participants with the session lasting from one to two hours. Krueger & Casey (2000) suggests between six and eight participants, as groups of this size show greater potential. According to Finch & Lewis (2003) and Rabiee (2007), focus groups consisting of between six and 10 participants encourage a good rapport, enabling interactive discussion to take place – the groups are large enough to obtain a variety of perspectives but not so large that they become unmanageable.

As can be seen in Table 3.2, the number of participants in each focus group in this thesis, ranged from four to 13.

Table 3.2: Size of focus groups

Group	Number of participants
Non-medical health workers	4
Nurses	5
GP practice	5
Public health	6
Retirees	7
Maori health provider	13

The Maori health provider group consisted of 13 members which is above the recommended number of participants. However, it was evident that the participants were used to discussing and debating issues relating to health care and that they felt comfortable doing so. Everyone contributed and no individual or splinter groups dominated the discussions.

3.4.3 Number of groups

In other studies, the number of focus groups used has ranged from a few groups to over 50 depending on the aims of the project, the resources available, the number of sub-groups affected by the research topic and whether other data collection techniques are being used as well (Kitzinger 1995, Lewis & Finch 2003). Kidd & Parshall (2000) considers that for the purposes of peer-reviewed social and health research, confidence in focus group findings is enhanced by conducting multiple groups and including other data sources.

The number of groups needed to fully explore a subject is usually determined by the 'saturation' principle (Kvale 1996). Saturation occurs when no new information emerges from group discussion, with subsequent groups only contributing repetitive information (Pickler 2007). Morse (1995) argues that "saturation is the key to excellent qualitative work" (p 147). It is the *richness* of the data that is important rather than the *quantity* of data gathered.

The six groups of participants were recruited from a cross-section of the population in order to obtain a wide variety of views. Although there was heterogeneity between the groups, and each group tended to focus on a specific aspect, as will be discussed later, overall the group discussions were very similar. It appeared that saturation had been reached as no new information was being elicited from the groups.

3.4.4 Facilitation

Focus groups are unique in their ability to generate data based on group interaction (Rabiee 2007). The main role of a facilitator is to ensure that the interaction is among the participants and not between the facilitator and participants. A good facilitator guides the discussion, asks probing questions, seeks clarification, refocuses the group when it strays off topic, and ensures that everyone has their say and that one or two participants do not dominate the group (Bender & Ewbank 1994, Brondani et al. 2008). A facilitator can adopt a more "interventionist style" by comparing and contrasting participants' views and by asking participants to clarify their opinions thereby encouraging group discussion (Kitzinger 1995, p 301).

The purpose of a focus group is to gain as much information and insight as possible rather than to obtain uniform answers or to reach a group consensus where *everyone* has to agree (Casey & Krueger 2004). A facilitator should be aware of 'censoring' or 'conformity' within

a group. 'Censoring' occurs when a participant withholds comments because they do not feel comfortable in the group or they are unsure how the data will be used, whereas 'conformity' refers to when a participant adjusts their comments in order to reach group consensus (Carey 1995).

To ensure that a good rapport is developed among participants, they need to feel at ease with the facilitator and the other group members. Meetings should be held in a setting where participants can easily see and hear each other (Casey & Krueger 2004). At the beginning of a meeting, the facilitator should greet the participants, discuss the purpose of the study, provide ground rules for discussion and explain how the data obtained from the focus groups will be used in the study.

When several focus groups are convened it is important to use the same format and the same materials for all groups to ensure consistency and comparability between the groups. At the end of each meeting the facilitator should present the issues raised at the meeting and seek confirmation or clarification from the participants (Kidd & Parshall 2000).

In this thesis each of the focus group meetings followed a similar format. Prior to attending a meeting, I sent all participants an email. The email contained a link to the online ranking survey (referred to in Section 3.3.1), an information sheet and confirmation of the time and place of the meeting.[79] I asked participants to do the online ranking survey before attending the meeting. Meetings were held at locations that were convenient for the participants. For example, the meetings for the GP practice, non-medical health workers and Maori health provider group were held in their workplaces. I provided refreshments but did not offer any compensation or incentives for attending the meetings. Meetings lasted from one to two hours.

I started each meeting by introducing myself and giving an overview of the research. I discussed some ground rules for the meeting and explained how the data from the focus groups would be used. Participants introduced themselves and signed a consent form before the results of the ranking exercise were discussed.[80] The meetings were recorded for me to transcribe later.

[79] A copy of the email sent to the focus group participants is in Appendix 3.2 and a copy of the information sheet is in Appendix 3.3.
[80] A copy of the consent form is in Appendix 3.4.

The results from the online ranking survey were used as a 'warm-up' for the meetings. When participants discovered that other participants had ranked the vignettes in a different order, discussion and debate followed. As participants discussed why they ranked particular treatments ahead of others, potential 'criteria' for the decision survey were suggested and I wrote these on a white-board.

Kitzinger (1994) used a similar approach in her study on AIDS and the media to initiate discussion in focus groups. In that study participants were given cards with statements relating to who might contract AIDS, and as a group, sort the cards into different piles according to the degree of risk. When participants disagreed with each other, the reasons for their choices became apparent.

At the end of each meeting participants reviewed the potential criteria (on the whiteboard) to ensure that they accurately reflected the group discussion and that nothing was missing. The meetings concluded with participants ranking, by consensus,[81] the 14 vignettes in terms of their overall benefit/value to society. A 'consensus' ranking was obtained so that the vignette rankings could be compared between the focus groups. This is discussed in Section 3.5.

3.4.5 Analysis

Analysing focus group data involves identifying and refining themes (Barbour 2005). Clearly documenting the analysis of qualitative data increases the rigour of a study by potentially allowing other researchers to verify the findings (Rabiee 2007).

The focus group meetings were recorded, with permission from the participants. Shortly after the meetings, the recordings were transcribed and summarised, and additional observations noted such as whether groups had dominant members. According to Rabiee (2007), although the main source of data analysis is the recorded discussions, being aware of non-verbal communication is important as it can add value to the data analysis. An advantage of me being both the facilitator and the transcriber was that I was aware of non-verbal communications such as raised eyebrows or nods of agreement.

[81] A consensus means that *most* of the participants in a focus group agreed with the vignette rankings. The 'consensus' differed depending on the number of participants in each group. For example, the 'consensus' for the nurses' group was three out of four whereas it was eight out of 13 for the Maori health provider group. The process used for arriving at a group consensus was for the group to agree on the four highest ranked vignettes and the four lowest ranked vignettes. This was usually quite easy. Agreeing on the ranks for the remaining six vignettes was more challenging.

Qualitative data can be analysed in a number of ways. For example, Computer-aided Qualitative Data Analysis software (CAQDAS), such as NVivo, is used to assist in transcribing, coding, interpreting and extracting the main concepts from qualitative data (Lewins & Silver 2007). Ritchie & Spencer (1994) suggests six key stages for analysing qualitative data: familiarisation, identifying a thematic framework, indexing, charting, mapping and interpretation. Krueger & Casey (2000) proposes an analytical framework encompassing frequency, specificity, emotions, extensiveness and big picture. The approach taken depends on how the qualitative data are collected and the topic being investigated.

It was not necessary to use CAQDAS or a formal framework in this thesis to identify and extract key concepts from the focus groups as the potential criteria were written on the whiteboard at each meeting and the participants were asked to confirm that the list was definitive. This is in contrast to other focus groups studies where the main themes are identified by a transcriber *after* the meeting. However, transcribing the six focus group meetings provided valuable information in terms of supporting statements and being able to compare the discussions across the groups.

Also, as the main purpose of the focus groups was to elicit potential criteria for inclusion in the decision survey, participants did not have to agree on how important the suggested criteria were, nor were they restricted with respect to the number of criteria they could suggest. This made the process relatively straightforward.[82]

3.5 Results of focus groups

A summary of the focus group meetings and a comparison of the ranked vignettes from the focus groups are presented in this section.

[82] When ranking the vignettes in a group, consensus was more difficult to reach. However, this exercise was done to compare the vignette ranks across the groups and was not the main objective of the focus group meetings.

3.5.1 Focus group summary

> "I would value anyone, for the public [sic] to do this kind of exercise, because it makes you think about things, because everyone gets on their own little bandwagon and wants it for their own and not for the good of all. People are thinking of other people".
>
> *Member of the retirees' focus group*

The main objective of having participants rank the vignettes before attending their focus group was to provoke thought to initiate conversation in the upcoming meeting. In this respect the ranking exercise was successful. Participants were already thinking about factors affecting health care prioritisation when they arrived at their meeting and when they discovered how other participants ranked the vignettes, the discussion flowed. Many participants found it difficult to rank the vignettes because there were so many factors to consider.[83] Some participants commented that it was easier to rank the treatments in a group because the group discussions helped them to clarify their thoughts and provided information that they had not previously considered.

Although there were a few dominant participants in the groups, all participants had the opportunity to offer their opinions.[84] Many participants said that they felt valued by being included in a focus group and that they were pleased to contribute to an area of research that they considered to be vital. Most participants acknowledged that health care prioritisation is a complex and difficult task but it is necessary, as our health budget cannot keep pace with the new treatments that are continuously becoming available.

Though cost was not included in the vignettes, some participants found it difficult to ignore the potential cost of treatment when ranking the vignettes and to just consider the benefits (refer to Figure 3.2).[85] This was particularly evident when participants had some idea of the relative magnitudes of treatment costs because they worked in the health area or had personal experience of the treatments.

---

[83] As discussed in Chapter 2, the cognitive difficulty of trading-off multiple criteria at one time is the reason why 1000Minds software and the PAPRIKA method will be used. (This is discussed further in Chapter 5.) Although it was difficult for some participants in the focus groups to rank the vignettes, the exercise was extremely useful in eliciting potential criteria for the decision survey.

[84] Dominant participants are those people who control the group by talking too much, strongly voicing their opinions or cutting out other participants who have an opposing view.

[85] As previously mentioned, participants were asked to ignore cost and to just consider the benefits/value of the treatment to society when ranking the vignettes. The reasons for excluding cost will be discussed further in Chapter 4 and the inclusion of cost alongside the benefits of treatment is explained in Chapter 9.

Some participants recounted personal stories regarding the treatments. However, by considering treatment *programmes* rather than individual patients, discussion could be focussed on societal benefit rather than individual benefit.

It was relatively easy to elicit 'potential criteria' for prioritising health care. When discussing why one treatment should receive priority over another, participants' comments illuminated potential criteria. For example, when participants were discussing 'growth hormone' in one of the meetings, two comments were: "but the person is young" and "we should treat young people first", thereby eliciting a potential criterion of 'age'. Although participants differed with respect to how strongly they felt about certain criteria, the objective was to elicit potential criteria for the decision survey, not to discover the strength of preference for those criteria (that comes later). This encouraged a good group dynamic – participants could share their thoughts without having to rigorously defend their point of view.

For example, when discussing 'statins for cardiovascular disease', 'lifestyle choice' was elicited as a potential criterion. Some participants felt that people who do not exercise and/or eat too much are personally responsible for their ill health and therefore should not have priority for treatment, whereas others felt that not everyone has the same control over their lives because of having a low income for instance and therefore should not be penalised. The groups agreed that 'lifestyle choice' was a factor in prioritising health care but participants were not required to agree on its relative importance in a prioritisation process. A summary of the meeting transcripts for each focus group is presented in Appendix 3.5.

It appeared that each group had a slightly different perspective as a result of the group membership: the nurses focussed on treating sick patients, regardless of the cause of illness; many of the retirees had personal experience of the treatments being discussed but nevertheless were able to focus on societal benefit; the participants in the non-medical health group emphasised the tension between caring for all members of society and patients' self responsibility; the GP practice accepted that not everyone could be treated because of a limited budget and therefore those who would benefit the most from being treated should receive priority; the public health group discussed the flow-on effects of patients not receiving treatment and the efficacy and delivery of treatment; and the Maori health provider group discussed building a stronger productive society by treating children, mothers, whanau and the community in a holistic way, regardless of ethnicity or lifestyle. Though each group tended to focus on a different aspect of prioritisation, the potential criteria elicited from each

group were similar overall. The list of potential criteria from each focus group meeting is presented in Table 3.3.[86]

Table 3.3: Potential criteria from focus groups

Nurses	Retirees	Non-medical health	Public Health	GP Practice	Maori Health Provider
number of people affected	number of people who would benefit	how many it treats	number of people (size of effect)	large number of people (large benefit impact across society)	numbers treated
preventative nature of treatment	preventative, education	preventative	ability to prevent	preventative	preventative
	early intervention			early detection	
the impact on family and society (medically, socially and economically)	enables people to stay in workforce, impact on family and society, economic impact	productivity, impact on family and society, economic impact	impact on family and others, society, flow-on effects	economic impact of not working, contribution to society, impact on family and society	impact on children, family, society, economic, productivity, flow-on effects
quality of life	quality of life	quality of life	quality of life	quality of life	quality of life
life expectancy (life extending)	extended life	length of life	length of life	life extending	length of life
effectiveness of treatment	success/ duration of treatment, strength of evidence	relative success	possible/proven efficacy/ effectiveness of treatment	success of treatment, length of treatment, impact on individual	
lifestyle choice	lifestyle choice, self responsibility	lifestyle		lifestyle choice	
age	age	age	age	age	age
health need (relative need and severity)		sickest	severity of illness	cure vs maintenance	need
are there alternative treatments?	are there alternatives?	life/death		alternative treatments or can pay for it themselves	
		obligations/ value to society	common good, increase equity (societal goal), ethnicity	mode and place of delivery	social need, cultural considerations: inequalities, accessibility, treatment of care, gender
	whether hereditary	religious/ political views			

[86] The means by which these suggested criteria were amalgamated into the criteria used in the decision survey is explained in Chapter 4.

3.5.2 Focus group rankings

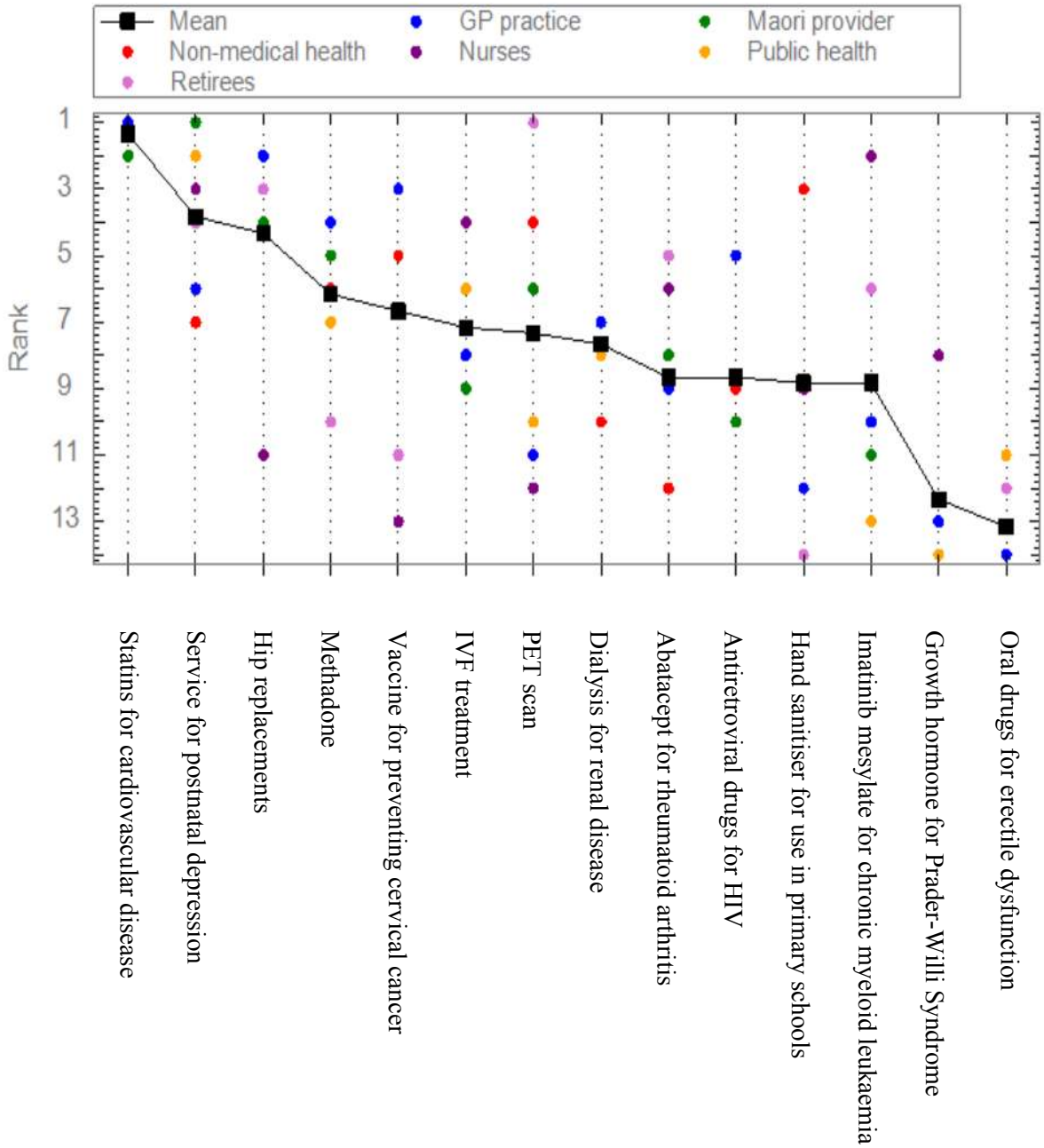
At the end of each focus group meeting the participants ranked the vignettes by consensus. The ranked vignettes for each focus group together with the mean and median ranks for all groups, are listed in Table 3.4 and presented graphically in Figure 3.4.

Table 3.4: Ranked health vignettes by focus group consensus

Health vignette	Nurses	Non-medical health workers	Retirees	Public health	GP practice	Maori health provider	**Mean***	**Median****
Statins for patients at high risk of cardiovascular disease	1	1	2	1	1	2	1.3	1.0
Service for postnatal depression	3	7	4	2	6	1	3.8	3.5
Hip replacements	11	2	3	4	2	4	4.3	3.5
Methadone for opioid addiction	5	6	10	7	4	5	6.2	5.5
Vaccine for preventing cervical cancer	13	5	11	5	3	3	6.7	5.0
IVF treatment	4	8	8	6	8	9	7.2	8.0
Positron emission tomography (PET Scan)	12	4	1	10	11	6	7.3	8.0
Dialysis for end-stage renal disease	7	10	7	8	7	7	7.7	7.0
Abatacept for last-line treatment of rheumatoid arthritis	6	12	5	12	9	8	8.7	8.5
Antiretroviral drugs for HIV	10	9	9	9	5	10	8.7	9.0
Imatinib mesylate for chronic myeloid leukaemia	2	11	6	13	10	11	8.8	10.5
Hand sanitiser use in primary schools	9	3	14	3	12	12	8.8	10.5
Growth hormone for Prader-Willi Syndrome	8	13	13	14	13	13	12.3	13.0
Oral drugs for erectile dysfunction	14	14	12	11	14	14	13.2	14.0

*The mean rank of each vignette is the average rank and is calculated by adding together all the ranks for that vignette and dividing by the total number of groups (i.e. six).
**The median value of each vignette is calculated by ordering the ranks for that vignette from lowest to highest and selecting the middle value.

Figure 3.4: Graph of vignette rankings by focus group



To assess the level of agreement between the groups in regard to the criteria rankings, Kendall's coefficient of concordance (or Kendall's W) was calculated. This statistic ranges from 0, showing no agreement between the groups, to 1, showing complete agreement between the groups.

As can be seen from Table 3.5, Kendall's W across the six focus groups is 0.553 at p=0.000 which indicates moderate agreement between the groups.[87] However, when vignettes that are ranked either relatively high or relatively low, are excluded from the rankings, Kendall's W is lower, showing less agreement among the groups. For example, when either the two highest-ranked vignette (i.e. statins and service for postnatal depression) *or* the two lowest-ranked vignette (i.e. growth hormone and oral drugs for erectile dysfunction) are excluded from the rankings, Kendall's W across the six focus groups falls to 0.418 (p=0.004) and 0.403 respectively (p=0.005). However, when the two highest-ranked vignettes (i.e. statins and service for postnatal depression) *and* the two lowest-ranked vignettes (i.e. growth hormone and oral drugs for erectile dysfunction) are excluded from the rankings, Kendall's W across the six focus groups falls to 0.179 (p=0.378) indicating little agreement among the groups. In contrast, when three of the middle-ranked vignettes (i.e. IVF, PET scans and dialysis) are excluded from the rankings, Kendall's W is 0.673 (p=0.000), and when six of the middle-ranked vignettes (see Table 3.4) are excluded from the rankings, Kendall's W is 0.693 (p=0.000).

Table 3.5: Kendall's coefficient of concordance

Rankings across the six focus groups	Kendall's W	p-value
14 vignettes	0.553	0.000
12 vignettes (2 highest ranked vignettes excluded)	0.418	0.004
12 vignettes (2 lowest ranked vignettes excluded)	0.403	0.005
10 vignettes (2 highest *and* 2 lowest vignettes excluded)	0.179	0.378
11 vignettes (3 middle ranked vignettes excluded)	0.673	0.000
8 vignettes (6 middle ranked vignettes excluded)	0.693	0.000
6 vignettes (4 highest *and* 4 lowest ranked vignettes excluded)	0.086	0.766

The ease with which participants ranked the vignettes by consensus is reflected in the value of Kendall's W. Each group started by agreeing on the 'top' four vignettes and the 'bottom' four vignettes. Though there was not total agreement within each group, it was relatively easy to reach consensus. However, ranking the other six vignettes was more difficult. As can be seen

[87] Kendall's coefficient of concordance is used when there are three or more sets of rankings. According to Field (2005, p 1011) significance tests for Kendall's coefficient of concordance are "relatively meaningless because the levels of agreement usually viewed as good in the social sciences are way above what would be required for significance".

in the Table 3.5, when the four highest-ranked vignettes *and* the four lowest-ranked vignettes are excluded from the rankings, there is very little agreement across the groups (W=0.086, p=0.766).

However, a significant value of W close to one, indicating a high level of agreement, might imply that the groups are using similar standards or criteria when ranking the treatments but does not necessarily mean that the ranking is correct or accurately reflects the preferences of the general public.[88]

Three of the vignettes were ranked almost identically. 'Statins for patients at risk of cardiovascular disease', was ranked either first or second by all of the groups. Although some participants felt that a patient's need for statins resulted from their lifestyle choice and therefore they should not receive priority, the fact that statins could help prevent over 66,000 heart attacks or coronary deaths led to its high ranking (see Figure 3.1). 'Growth hormone treatment for Prader-Willi Syndrome' was ranked low. Although the treatment benefits children which influenced some groups' rankings, it treats only one aspect of the Syndrome (see Figure 3.1). After some debate, 'oral drugs for erectile dysfunction' was ranked last by four of the six groups and third or fourth last by the other two groups. Relative to the other 13 treatments, the value or benefit to society of this treatment, was considered to be very low.

Some treatments, particularly those associated with a poor lifestyle choice (e.g. methadone), provoked the most debate. Some participants argued that the negative flow-on effect to families and/or society of not treating a patient with methadone is wide-reaching whereas other participants felt that because it was a poor lifestyle choice that caused the need for treatment these patients should not receive priority. Similarly with HIV – some respondents felt that if a patient contracted HIV by accidentally getting pricked with a contaminated needle or a baby contracted HIV through breastfeeding for instance, then they should receive priority. But if the person contracted HIV by having 'unsafe' sex, then they should not receive priority. Participants were encouraged to think of a patient group comprised of a range of individuals and/or causes of illness, rather than particular individuals when discussing the health treatments.

'Hand sanitiser' was the most difficult treatment for the groups to rank as it tended to polarise participants. Some participants thought that hand sanitiser should receive a low priority

[88] The vignettes are also ranked later by using the criteria weights from the decision survey. This is discussed in Chapter 6.

because washing hands with soap and water is a viable alternative whereas others thought that hand sanitiser should receive a high priority because of its positive flow-on effects, preventing illness spreading to a large number of people.

As can be seen from the above examples, using a wide range of vignettes (some of which could be considered controversial such as methadone) helps to identify the relevant (often conflicting) criteria. For example, statins (individual benefit, numbers treated), growth hormone (individual benefit, age of patient), methadone (lifestyle choice, societal benefit), hand sanitiser (alternative treatment available, societal benefit) and so on.

To assess the usefulness of the focus groups' data, the methodology needs to be examined. This is discussed in the next section.

3.6 Validity of focus group research

In quantitative research the concepts of 'validity' and 'reliability' are commonly used to assess the robustness and quality of the research. 'Validity' measures how well a study captures what it intends to measure and 'reliability' measures whether the findings from the research are replicable in similar situations. [89]

Morse (2002) argues that 'validity' and 'reliability' can also be used to assess *qualitative* research if verification strategies are implemented throughout the research to shape and direct it. Carey (1995) also considers 'validity' and 'reliability' to be useful for qualitative research, with the comparable concepts being 'credibility' and 'usefulness'.

Mays & Pope (2000) considers that qualitative research cannot be judged by the conventional methods of 'validity', 'reliability' and 'generalisability'[90] used in quantitative studies but instead the broad concepts of 'validity' and 'relevance' can be used in a different way to meet the goals of qualitative research.

Lincoln & Guba (1985) suggests four criteria for evaluating the trustworthiness of a qualitative study: 'credibility', 'transferability', 'dependability' and 'confirmability'. Though

[89] There are many different validity and reliability measures. For example, construct validity, content validity, face validity, criterion validity, predictive validity, internal and external validity, inter-rate reliability, test-retest reliability (which is discussed in Chapter 6), internal consistency and inter-method reliability. Which measure(s) are used depends on a variety of factors including the type of study (qualitative or quantitative) and the survey design.

[90] Generalisability measures the extent to which the research findings can be applied to the wider population.

these criteria are different from those used in quantitative research, they have parallels with validity, reliability and generalisability.

'Credibility' refers to the internal validity or 'truth value' of a study; that is, how credible the results are from a participant's point of view. This can be assessed by 'member checking', which according to Guba (1981) is the most crucial technique for establishing credibility. With 'member checking', the data and/or interpretations are presented to the participants in the study so that they can confirm the credibility of the information.

'Transferability' refers to how well the findings of the study can be generalised to other settings. 'Dependability' refers to the consistency or the reliability of the findings. A study would be considered dependable if similar results were produced in a repeat of the study. The degree to which the findings of the study are able to be confirmed or corroborated by others is 'confirmability'. The study should be driven by participant interaction and not by the motivation or interest of the researcher.

In addition, an audit trail that clearly documents the data collection methods and analysis should be established. This will validate whether the researcher's interpretation of the study is an accurate representation of the data (Mays & Pope 2000).

The concepts of 'credibility', 'transferability', 'dependability' and 'confirmability', together with an audit trail, were used to evaluate the robustness of the data obtained from the focus groups in this study.[91]

'Member checking' was used to establish the credibility of the potential criteria. As mentioned earlier, the vignettes were verified by health experts for accuracy and a pilot group ranked the vignettes to test the ranking survey and ensure that the vignettes were comprehensible. When the participants in the focus groups discussed the vignettes, potential criteria were written on a whiteboard. At the end of the meeting participants were asked to check whether the criteria accurately reflected the groups' discussions and that the list was exhaustive.

[91] These concepts are most often used when the *main* methodology is qualitative research. In this thesis, focus groups are used for *exploratory* research prior to quantitative research. However, it is still useful to use these concepts to evaluate the validity and reliability of the focus group research.

As discussed in Section 3.4.5, conducting further focus groups or reconvening the same focus groups was unlikely to generate additional criteria to those already obtained and therefore I concluded that 'saturation' had been reached. Wensing & Elwyn (2003) suggests a different mix of participants is unlikely to yield exactly the same information. However, the criteria elicited from each group were similar even although the composition of the groups differed.

A literature review (conducted before the focus groups were established) reveals that the potential criteria elicited from the focus groups in this thesis closely align with the criteria elicited in other studies.[92] This suggests that the criteria elicited from the focus groups, with some modification to meet the needs of the survey, can justifiably be included in a decision survey which will be undertaken by a wider cross-section of the general population.

Finally, in terms of 'confirmability', by establishing a clear audit trail other researchers can examine the focus group research to ascertain the legitimacy of the findings. Every focus group meeting followed the same format with participants discussing the vignettes, agreeing on potential criteria and ranking the vignettes by consensus. To ensure consistency, I conducted, facilitated and transcribed all focus group meetings.

3.7 Conclusion

In this chapter the process of eliciting potential criteria for use in the decision survey is described. Six focus groups were established with members representing a cross-section of the general public. Before participants attended a focus group meeting they completed an online ranking survey where they were asked to rank a list of health treatments in terms of their value/benefit to society. The ranking exercise formed the basis of discussion in the focus groups. From the group discussions potential criteria were elicited for the decision survey (discussed in Chapter 5).

The process of amalgamating the potential criteria from the focus groups with suggested criteria from health experts and the literature is described in the next chapter.

[92] A literature review of other studies eliciting criteria relating to health care prioritisation is included in Chapter 4.

Appendix 3.1: Email sent to pilot group participants for ranking exercise

Hi {fullname}

As part of my PhD research on prioritising healthcare services, I have developed a survey and would be very grateful if you would take part.

To start, click on the link below. There you'll be asked to rank 14 health 'technologies' (ie treatments) according to what ***YOU*** think are their **overall benefits/value to society**.

What matters is ***YOUR*** personal opinion - there are no 'right' answers!

All the info you need is in the link below, but in case it's useful to you, a copy of the treatment descriptions has been sent to you in a separate email.

(*Please do not forward this email to anyone else, as each survey can be completed by 1 person only*.)

Click on this link to begin:
{url}

I'd also appreciate any feedback: trudy.sullivan@otago.ac.nz or ph 479 8134

Many thanks for your help.

Trudy.

Appendix 3.2: Email sent to focus group participants for ranking exercise

Hi {fullname}

Thanks for agreeing to be involved in my study on how to prioritise healthcare treatments.

As a warm-up exercise prior to our meeting on xxx, I'd like you to rank some health 'technologies' (i.e. treatments). To do this click on the link below. There you'll be asked to rank 14 health technologies according to what ***YOU*** think are their overall **benefits/value to society**.

What matters is ***YOUR*** personal opinion - there are no 'right' answers! So don't stress or take too much time to do it.

The treatment descriptions are included in the link below, but incase it's useful, I've sent a copy of them to you in a separate email.

(*Please do not forward this email to anyone else, as each survey can be completed by 1 person only*.)

Click on this link to begin:
{url}

If you have any questions and/or comments my contact details are: {reply-to} or ph 479 8134 (w) 489 1148 (h)

See you xxx at xxx.

Trudy

Appendix 3.3: Information sheet for focus group participants

INFORMATION SHEET FOR PARTICIPANTS

HEALTH TECHNOLOGY PRIORITISATION: setting priorities in a publicly-funded health system

Background/Motivation

New Zealand, like many other countries, is faced with the problem of allocating limited healthcare resources among different health and disability support services. To ensure our health dollars are being spent wisely, health care decision makers need to adopt robust, transparent processes for setting priorities, preferably with the involvement of key stakeholders (essentially anyone who is involved in health – patients, health professionals, taxpayers, policy makers etc).

The focus of my PhD is to discover the relevant criteria (and their importance) for setting priorities in a publicly-funded health care system, by surveying the key stakeholders.

What will participants be asked to do?

Should you agree to take part in this voluntary exercise, you will be asked to complete a simple ranking exercise before coming to a workshop. I will send you an email with a link to a survey where you will be asked to rank 14 health treatments in terms of their overall benefit to society. At the workshop we will discuss the type of things you thought about when ranking the health treatments and then we will try to group these into overall headings, i.e. define some criteria. Then we will rank the health treatments by consensus. The workshop should take approximately one hour.

What is the purpose of the workshop?

To come up with some criteria that could be used in a survey to elicit preferences. Specifically a computer survey will be sent to a large sample group (you can be included in that too if you wish) to determine the relative importance of each criteria (i.e. find out people's preferences). A demographic survey will be attached to the main survey so the preferences of key stakeholders can be analysed according to various characteristics.

Questions?

If you have any questions about our project, either now or in the future, please feel free to contact either:

Trudy Sullivan
Department of Economics
University Telephone Number: 479 8134

or

Associate Professor Paul Hansen
Department of Economics
University Telephone Number: 479 8457

Appendix 3.4: Consent form for focus group participants

CONSENT FORM FOR PARTICIPANTS

HEALTH TECHNOLOGY PRIORITISATION: setting priorities in a publicly-funded health system

I have read the Information Sheet concerning this project and understand what it is about. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage.

I know that:-

1. My participation in the project is entirely voluntary;

2. I am free to withdraw from the project at any time without any disadvantage;

3. This project involves an open-questioning technique where the precise nature of the questions which will be asked have not been determined in advance, but will depend on the way in which the discussion develops and that in the event that the discussion develops in such a way that I feel hesitant or uncomfortable I may decline to answer any particular question(s) and/or may withdraw from the project without any disadvantage of any kind;

4. The results of the project may be published and available in the University of Otago Library (Dunedin, New Zealand) but every attempt will be made to preserve my anonymity.

I agree to take part in this project.

....

(Signature of participant) (Date)

Appendix 3.5: Summary of focus group meetings

Group: **NURSES**

Composition: Five female participants: two registered nurses working in a private hospital and three registered nurses working in a public hospital.

Ranking: Individual ranking exercise prior to meeting: 3 participants
Majority consensus for group ranking exercise: 3 out of 5

Discussion: All New Zealand residents should be treated equally regardless of ethnicity. The suggested criteria should encompass the Treaty of Waitangi, considerations such as higher health need for Maori, family focus and self responsibility.

"Socially disadvantaged" was not considered to be a factor. If people are being treated based on need it shouldn't matter whether they are socially disadvantaged or not. Possibly the more socially disadvantaged a person is, the higher their need and therefore they are treated on that basis.

It was apparent that participants found it difficult to separate their personal preferences from social benefit. For example, for those nurses who worked in particular areas or had personal experience of a specific treatment, it was difficult for them not to rank those treatments high even if they felt that other treatments may have a higher social benefit.

The discussion was fairly muted – there were no major disagreements or robust discussions possibly because the nurses worked in both the private and public sectors and they did not want to make comparisons. A person's poor lifestyle was not considered to be a criterion. One nurse suggested that this was because of their training – that all people should be treated equally and that why or how a patient seeks treatment should not be considered.

Draft criteria: The number of people affected, preventative nature of treatment, the impact on family, the impact on society (medically, socially and economically), quality of life, life expectancy (life extending), effectiveness of treatment, age, health need.

Appendix 3.5: Summary of focus group meetings

Group: **NON-MEDICAL HEALTH CARE WORKERS**

Composition: Four participants (three female, one male) working in non-medical roles in the health sector.

Ranking: Individual ranking exercise prior to meeting: 4 participants
Majority consensus for group ranking exercise: 3 out of 4

Discussion: Cost was difficult to keep out of the discussion. Whether people could pay for the treatment themselves was a consideration as well as whether the treatment was a cure or whether it would lead to further on-going costs.

Because of the society we live in, our obligation as a caring nation needs to be considered. For example, instead of giving low priority to a methadone programme for drug users, the flow-on effects to family and society should be considered. There is tension between acting as a caring society and the self-responsibility of its citizens.

The political influence on health budget allocation was discussed. The group felt that health treatments or programmes were sometimes given more or less prioritisation based on a political agenda, lobby groups etc. Religious views may also influence prioritisation decisions, for example HIV treatment.

A downside to having newer and more available treatments is that for some people (eg the elderly) we are taking away the choice of how they may wish to die. We are prolonging life, but at what quality?

Draft criteria: Relative success of treatment; preventative; numbers treated; alternative treatment; quality of life/length of life; productivity; impact on family; impact on society; economic impact; impact on health budget; lifestyle; age; life/death; sickest; obligations/value to society; religious/political views.

Appendix 3.5: Summary of focus group meetings

Group: **RETIREES**

Composition: Seven participants (four females, three males), retired, over 65 years.

Ranking: Individual ranking exercise prior to meeting: 7 participants
Majority consensus for group ranking exercise: 4 out of 7

Discussion: Although many of the participants in the group had personal experience of the treatments that they were asked to rank, they were able to look at the bigger picture and think of society rather than their personal preferences.

Participants discussed how people create their own problems with less exercise, eating more, lower family values, use of drugs, dependency on the welfare system, and that there is a flow-on effect to family and friends. There needs to be more self-responsibility.

People now have to wait or miss out on treatments whereas it used to be the case that most treatments were freely available.

It was suggested that more should be spent on researching the causes of disease rather than the treatments themselves and to explore alternative treatments, emphasise education and prevention with a view to keeping people in the workforce. While some treatments such as IVF, help a couple and their family, they are not helping people who have a low quality of life or reduced length of life.

Draft criteria: Early intervention (cost effective); preventative (education); enables people to stay in the workforce; extended life; number of people who would benefit; quality of life; impact on family/society; success/duration of treatment; lifestyle choice; age (economic impact of not working); self-responsibility; alternative treatment; hereditary factors; research/strength of evidence.

Appendix 3.5: Summary of focus group meetings

Group: **PUBLIC HEALTH AND PREVENTIVE AND SOCIAL MEDICINE**

Composition: Six participants (4 females, 2 males), one from Plunket, two from Public South Health, three from the University of Otago Preventive and Social Medicine department.

Ranking: Individual ranking exercise prior to meeting: 6 participants
Majority consensus for group ranking exercise: 4 out of 6

Discussion: The flow-on effect of not receiving treatment was considered important – the impact on children, the family and wider society now and in the future; for example, decreased quality of life, possible increase in crime, effect on mental wellbeing and relationship problems.

Improving delivery of health care would help to increase societal equity, particularly for Maori who have a higher health need.

How well informed the public are about certain treatments was also discussed. There is a big difference between informed lay people and uninformed lay people so it is important that people receive information, through health promotion for instance, to know why health professionals make the decisions they do.

It was felt that proven efficacy and effectiveness of treatments is necessary before a treatment is included in a prioritisation process.

Draft criteria: Age, impact on family and others, impact on society, quality of life (physical, mental, social, spiritual), possible/proven efficacy/effectiveness of treatment; ability to prevent; number of people (size of effect); flow-on effects; severity of illness; length of life; principles (common good, increase in equity (societal goal), ethnicity).

Appendix 3.5: Summary of focus group meetings

Group: **GP PRACTICE**

Composition: Five participants (four female, one male) – two GPs, one GP registrar, one practice nurse, one practice manager.

Ranking: Individual ranking exercise prior to meeting: 5 participants
Majority consensus for group ranking exercise: 3 out of 5

Discussion: Treating sick people on a daily basis and being very familiar with having to ration health care heavily influenced this group's discussion and ranking of health treatments. They looked at the ranking exercise clinically and pragmatically, while appreciating the impact on individuals, family and society.

They emphasised the costs and benefits of treatment in terms of lost productivity, ongoing costs of treatment and treating younger people before the elderly because of the greater potential younger people have to contribute to society.

Discussion was held around the mode of delivery (hospital vs community care) and whether treatment is a cure or just maintenance (ie does a treatment fix a person or will they need treatment for life?).

The group acknowledged that not everyone can be saved and decisions have to be made, albeit very difficult decisions. It is tragic at a personal level but someone has to make the decisions.

Discussion was at times dominated by one group member but group majority consensus was still able to be reached.

Draft Criteria: Economic impact of not working (contribution to society); alternative treatments (also affordability – if low cost can pay for it themselves); large number of people (large benefit impact across society); early detection; quality of life; preventative; lifestyle choice; age; impact on society/family; success of treatment – length of treatment; impact on individual; mode and place of delivery; cure vs maintenance; life extending.

Appendix 3.5: Summary of focus group meetings

Group: **MAORI HEALTH PROVIDER**

Composition: 13 participants (9 female, 4 male) – working in the areas of childrens' health and education; children 1-3 years and their whanau; road safety and HPV for teenagers; cardiovascular disease and diabetes; alcohol and drug; and healthy lifestyles.

Ranking: Individual ranking exercise prior to meeting: 8 participants
Majority consensus for group ranking exercise: 8 out of 13

Discussion: A holistic view of health needs to be considered as the flow-on effects impact on the family and wider community. Emphasis needs to be placed on children, mothers, family, mental health, justice, the community and building a stronger more productive society.

A person's lifestyle should not be considered as everyone has different resources, different backgrounds, different strengths and weaknesses.

Health problems such as cardiovascular disease and diabetes for example, are becoming more predominant. Research needs to be done on prevention and health promotion as well as treatment to help stem the flow of the people who are coming through with health problems.

People should be treated based on need regardless of race. However, the current delivery of health services and access to treatment is causing inequalities and disparities in health amongst Maori. Practitioners of health need to work collaboratively to ensure the appropriate delivery of, and access to, health services.

Criteria: Impact on children, family, society (including economic); flow-on effects; preventative; quality of life; length of life; social need; need; numbers treated; cultural considerations (inequalities, accessibility, treatment of care, age, gender).

~ Chapter 4 ~

Establishing criteria for the decision survey

4.1 Introduction

> "There are limited resources and if you have got to choose between a 35-year old and a 65-year old, as a health professional I would have to say you choose the 35-year old and then you make it all complicated by saying that the 35-year old is a drug user who smokes and the 65-year old doesn't and cares for two grandchildren. It's not cut and dried."
>
> *Member of the GP practice focus group*

In this chapter the process of amalgamating the suggested criteria from the focus groups with criteria from the literature and other relevant information is explained. The chapter begins with a review of studies exploring prioritisation of health care. The potential criteria elicited from the focus groups are discussed alongside comparable criteria from the literature. The chapter concludes with the criteria and levels to be used in the decision survey.

4.2 Literature review

As discussed in Chapter 2, there are two main theoretical approaches to explicitly prioritising health care services (Logan et al. 2004, Sabik & Lie 2008). An 'institutional' approach which uses principles to guide prioritisation and a 'technical' approach which uses specific criteria or judgements to decide which health services should be publicly-provided. In this section a range of studies investigating the principles that underpin health care prioritisation are discussed together with several studies that explore the specific criteria that could be included in a health prioritisation process.

4.2.1 Principles underpinning health technology prioritisation

According to Daniels & Sabin (1998) priority-setting is considered to be legitimate and fair if the decision-making process is transparent and based on reasons that 'fair-minded' people understand and consider reasonable. The authors propose a framework referred to as

'accountability for reasonableness', in which clinicians and patients become part of the rationing debate. Four conditions are suggested in order to ensure a fair priority-setting process: publicity, relevance, appeals and enforcement. The publicity condition requires that decisions and rationales relating to the funding of new technologies are publicly accessible. The relevance condition ensures that the people involved in the decision-making process agree that the evidence, reasons and principles used to prioritise services are relevant to meet the needs of the population. The appeals condition allows for decisions to be challenged, particularly when new evidence becomes available. The enforcement condition ensures the first three conditions are met through either voluntary or public regulation.

In a follow-up paper in 2008, Daniels & Sabin (2008) found that in a wide range of contexts, decision-makers are explicit when making rationing decisions, they obtain buy-in from relevant stakeholders, and they revise their decisions when new evidence becomes available, which are the three key elements of 'accountability for reasonableness'.

Dolan et al. (1999) used focus groups to evaluate public support for ethical principles of health care rationing. The participants were given a hypothetical rationing choice involving four patients. The ensuing discussion elicited some general ethical principles which were then compared with rationing principles from the theoretical literature.[93] The groups favoured three main rationing principles: a broad 'rule of rescue' giving priority to those in immediate need, health maximisation, and equalisation of life time health.

When investigating the allocation of scarce medical interventions such as organ transplants and vaccines, Presad et al. (2009) recommend a multi-principle allocation system because no single principle can incorporate all "morally relevant values" (p 423). The authors evaluated three multi-principle systems – QALYs, disability adjusted life years and the United Network for Organ Sharing points systems (in the US) – and argue that these systems do not adequately address the importance of fair distribution. An alternative multi-principle system is suggested by the authors: the 'complete lives system'. Core ethical values such as prioritising adolescents and young adults over infants, considering an individual's capacity to live a 'complete' life, making choices between equally ill recipients, and prioritising individuals to

[93] The five main rationing principles identified by Cookson & Dolan (1999) from the theoretical literature are: 'lottery' principles (e.g. first-in, first-served), distribution according to immediate need (i.e. 'rule of rescue'), health maximisation (distribute health care to increase the aggregate health of the whole community), equalising lifetime health or 'fair innings' (health care should be distributed to reduce health inequalities) and equalising opportunity for health or 'choicism' (priority should be given to those who suffer ill-health through no fault of their own).

enable or encourage future usefulness form the basis of the 'complete lives system'. The system consists of four overarching principles – treat people equally, favour the worst-off, maximise total benefits, and promote and rewarding social usefulness.[94]

In the studies just mentioned, various principles have been suggested to guide health care prioritisation. Two of the conditions suggested by Daniels & Sabin (2008) within their 'accountability for reasonableness' framework, 'publicity' and 'relevance', are fulfilled when using MCDA because of its transparent and inclusive approach (Baltussen 2006). However, when prioritising health care treatments and/or patients using principles instead of explicit criteria, it can be difficult to ensure consistency and transparency. Nevertheless, the principles can be used to develop specific criteria within a prioritisation framework.

Golan et al. (2010) reviewed the criteria used for HTA in 11 countries and the US State of Oregon. The authors found that the criteria used for prioritising technologies fall into three main groups "consistent with the main principles of allocative justice": "(a) need, appropriateness and clinical benefits; (b) efficiency and (c) equality, solidarity and other ethical or social values" (p 129). Table 3.1 lists the principles of allocative justice and the criteria used by international bodies to prioritise new health technologies. All the criteria in Table 3.1 were mentioned by participants in the focus groups.

[94] For example, prioritising doctors who develop vaccines or people who agree to improve their health so that they will not use so many health resources in future. The difficulty is agreeing on who would be 'useful'.

Table 4.1: Main criteria and 'other' considerations used internationally for prioritising new technologies

Principles of allocative justice	Criteria
Need	General Severity of condition Availability of alternatives
Appropriateness	Efficacy and safety Effectiveness
Clinical benefits	General Effect on mortality (life saving) Effect on longevity Effect on health-related quality-of-life
Efficiency	Cost-effectiveness/benefit Budgetary impact Cost
Equality	General Accessibility to the service Affordability to the individual
Solidarity	
Other ethical or social values	Autonomy Public health value Impact on future generations
'Other' considerations	
Quality of the clinical and economic evidence	
Other considerations not elsewhere classified	Strategic issues consistent with previous decisions and precedents

Source: Golan et al. (2010)

Provision of publicly-funded health care affects the general public and therefore, as discussed in Chapter 2, the general public or key stakeholders should be involved in the priority-setting process. Several studies that have included key stakeholders in exploring and/or defining specific criteria to be used in prioritising health care are presented below. Each study uses a different method to elicit potential criteria from the participants.

4.2.2 Criteria for rationing health services

One of the earliest studies involving the public in prioritising health services was a study by Bowling (1996). A representative sample of the British population was asked to rank 12

health care treatments covering a wide range of health problems. They were also asked questions relating to priority-setting and budget allocation. The highest priority (ranked first) was given to children with life-threatening illnesses followed by treatment for people who are dying. The lowest priority was given to infertility treatment and treatments for people aged 75 and over with life-threatening illnesses.

Stronks et al. (1997) conducted a series of six workshops involving a cross-section of the general public in The Netherlands. Each panel was asked to play the role of a parliamentary committee charged with selecting services (from a list of 10) that could be funded from a reduced health budget. The panel was presented with selection criteria proposed by the Dunning report[95] and information on the 10 treatments. The authors found that there was substantial variation in the way the groups prioritised the services. GPs considered access to services for those in need and inability to pay; specialists considered prevention, caring for the sick, and age; patients considered chronic disease and acute care; the public considered the financial resources of an individual and individual responsibility; and health insurers considered health risk and inevitable health needs.

Menon & Stafinski (2008) conducted a citizens' jury to develop criteria for setting priorities in HTA in Canada. Expert witnesses, including patients, policy-makers and clinicians, gave presentations to the jury. Participants took part in a series of priority-setting exercises using mini technology scenarios. The jury identified 10 criteria which should be considered when evaluating new technologies. One criterion – cost – was subsequently excluded from the list as the jury felt that considering *per-patient* cost alone, does not provide enough information to assess the overall budget impact of funding a new technology. The nine criteria, listed in order of importance, were: 'potential to benefit a number of people', 'potential to extend life with quality', 'potential to improve quality of life', 'potential clinical benefit over existing treatment(s)', 'lack of an alternative', 'potential to detect a condition which, if treated early, averts future costs', 'potential for additional applications', 'potential to extend life', and 'completeness of data on adverse affects'.

[95] In 1997 in The Netherlands, a Government Committee was commissioned to write a report, that became known as the Dunning report (after the Committee's chairperson), relating to prioritisation in health service provision. The Dunning report suggested that publicly-funded health care services should meet the following criteria: necessary care (care which is necessary to maintain or restore health), effectiveness (treatment has to be proven and documented), efficiency (efficient delivery based on the results of cost-effectiveness studies) and individual responsibility (when an individual cannot afford to pay for treatment).

Golan et al. (2010) used a conjoint analysis survey[96] to determine the relative weights of criteria that could be used in a prioritisation process. (The same software and scoring methodology used by Golan et al. is used in this thesis, i.e. 1000Minds and the PAPRIKA method, explained in Chapter 2.) The criteria were 'lives saved', 'life-prolongation benefits', 'quality of life gains', 'if this technology were not to be funded' and 'other important social/ethical benefits, e.g. targeted to children/minorities; reduces health gaps etc'.

To ensure that all possible considerations are incorporated into the criteria for the decision survey in this thesis, potential principles and criteria garnered from the literature are considered alongside the potential criteria from the focus groups and comments from health experts. This is discussed in the next section.

4.3 Developing the criteria for the decision survey

> "The criteria are the measures of performance by which the options will be judged, and must be carefully selected, to assure completeness, feasibility, and mutual independence, and avoid redundancy and an excessive number of criteria."
>
> *Baltussen & Niessen (2006, p 4)*

To ensure that the criteria for the decision survey encompass as many considerations as possible relating to health prioritisation, criteria from the literature (discussed in the previous section), criteria suggested from the focus groups and comments from health experts were combined and sorted into groups based on similarity. The potential criteria are listed in Table 4.2 (the points under each heading relate to suggested criteria from the focus groups, the literature or health experts).

[96] The same software and scoring methodology Golan et al (2010) used to determine the relative criteria weights is used in this thesis, i.e.1000Minds software and the PAPRIKA method.

Table 4.2: Potential criteria for the decision survey

Age • fair innings • priority for which group: babies, children, adolescents, elderly • targeted to young people
Alternative treatments • is there an alternative treatment available? • could the patient pay for it themselves? • rare diseases, orphan drugs
Equality/equity/access • ethnicity • socioeconomic, reduce disparities, poverty, vulnerable • economic welfare • distribution, geography
Need • urgency, sickest, worst-off, severity, high health need • rule of rescue • life, death, life saving
Quality of life (HRQoL) • potential for improvement in quality of life including physical, mental, social and spiritual aspects
Length of life • extending life • size of benefit
Societal benefit • impact on family • impact on society • ability to continue working, productivity, impact on economy, economic growth • spill-over effects, externalities, adults with children
Preventative • early intervention • prevents more harm in the future
Lifestyle choice
Effectiveness • success of treatment • duration of treatment, side effects • short-term vs long-term benefits, cure vs maintenance • comfort care, palliative care
Number of people affected • large gains for small number vs small gains for large number

As mentioned in Chapter 3, though cost is a potential criterion, it will not be included in the decision survey but will be considered alongside the benefits of treatments. The reasons for excluding cost are discussed in Section 4.4.

The 11 potential criteria were amalgamated into six criteria for the decision survey. In the following sub-sections, the 11 potential criteria are discussed, and the final six criteria and their associated levels are presented.

4.3.1 Age

Age was raised as a potential criterion in every focus group.

> "Age comes into it [health care prioritisation] for me hugely. It matters to the individual but it also matters to society. A five-year old has got the potential to contribute more to society than someone who is 70. I know that is ageist."
>
> *Member of the GP practice focus group*

> "Giving a two-year old 10 years of life is quite different from giving a 60 year-old 10 years of life. It's not just about the sickest, it's about age."
>
> *Member of the non-medical health workers' focus group*

> "Nowadays with grandparents looking after children it might be very important to have those 70 to 75-year olds."
>
> *Member of the Maori health provider focus group*

The fair innings argument (FIA) espoused by Harris (1985) and Williams (1997) is "that everyone is entitled to some 'normal' span of health (usually expressed in life years) ... and anyone failing to achieve this has been cheated, whilst anyone getting more than this is 'living on borrowed time'" (Williams 1997, p 117). In other words when people reach a certain age, they have had a 'fair innings' and their expectations for health care should not be the same as when they were young. This argument of equalising lifetime health is used to justify moving resources from the elderly to the young.

Nord et al. (1996) distinguishes between egalitarian ageism (the right for an individual to enjoy additional life years the fewer life years they have already had) and utilitarian ageism

(young people derive greater health benefits due to their greater life expectancy). The authors found that there was some support for including age in a prioritisation process.

Persad et al. (2009) considers that the youngest members of society have a stronger claim to life-saving resources because they have had the fewest life years and that this is not ageist because "treating people of different ages differently does not mean that we are treating persons unequally" as everyone ages (p 425). However, Persad et al. (2009) does not advocate a strict youngest-first policy as this would allocate resources to babies before adolescents. The authors contend that adolescents are more deserving of health care than babies because adolescents have already received a greater investment in their future compared with babies.

Rivlin (2000) disputes that the FIA can be used to support rationing of health care by age. The author has a number of concerns principally related to the definition and use of 'fairness'. He questions the notion of a 'fair share of life' and whether we can expect a fair share of resources when it is difficult to establish the exact amount of resources to be shared, and over which time period. He believes a major problem with the FIA is that fairness is limited to length of life and does not account for other considerations such as quality of life or the cause of illness in terms of a person's lifestyle choice. Rivlin (2000) does not believe that the rationale for the FIA is persuasive and that it may lead to discrimination against the elderly.

Farrant (2009) also argues against applying the FIA to justify redistributing health resources. How long we expect to live (a normal life span) is a combination of both biology and human intervention. As biomedical technologies continue to be developed and applied, the normal life span increases which further exacerbates the problem of fairly redistributing health care resources.

Baltussen et al. (2006) found that in developing countries preference is often given to adults for economic reasons as they are more productive than children and contribute more to society. It is often this argument that older people use to justify their 'entitlement' to health care.

Whether or not the FIA justifies redistributing health care resources, age is not usually considered in health care prioritisation on the grounds of discrimination. Kapiriri & Norheim (2004) explored Ugandan stakeholders' acceptance of criteria for setting priorities in their health care system. The Ugandan stakeholders considered age to be an unacceptable criterion

as well as race, religion, gender, social power and influence, mental features, lifestyle, genetic background and sexual orientation.

Diederich et al. (2011) used a mixed-methods approach to explore priority-setting in health care in Germany. A qualitative interview study was used to determine the considerations underlying priority-setting decisions. As 'age' appeared to be the most controversial consideration, a quantitative survey was undertaken, using a variety of questionnaire items including age-related questions, health care scenarios and discrete choice alternatives, to determine whether age should be considered in priority-setting. The authors found that there was no clear support for using age as a prioritisation criterion.

When developing guidance for the National Health Service in the UK, NICE's Citizens' Council acknowledged the debate over age discrimination in the allocation of health care resources and advised that health should not be valued more highly in some age groups than in others.

The New Zealand Human Rights Act 1993 makes it unlawful to discriminate on the basis of age. However, in practice age is already a prioritising factor, albeit in terms of clinical efficacy, in the allocation of some health care services in New Zealand. For instance, women over 40 years are ineligible for *in vitro* fertilisation funding, women aged between 45 to 69 years are eligible for free breast-screening and certain immunisations are offered free to babies (e.g. diphtheria/tetanus/whooping cough), children (e.g. tetanus), adolescents (e.g. human papillomavirus) and adults (e.g. flu injections for adults 65 years and over).

However, in various surveys to elicit people's attitudes towards prioritisation, the predominant view is that the young should be given priority over the old (Williams 1997). For example, in 1996 2005 people in the UK were interviewed to elicit their views on priorities for health services (Bowling 1996). The highest priority was given to "treatments for children with life-threatening illnesses" and the lowest priority was given to "treatment for people aged 75 and over with life-threatening illness" (Bowling 1996, p 670).

Although age as a criterion for priority-setting may not be politically viable, it was suggested as a potential criterion in every focus group and therefore it is included as a criterion in the decision survey.

As mentioned in Chapter 2, each criterion in the decision survey includes levels which are ranked from the lowest to the highest. In most cases the levels for a criterion are inherently ranked. For example, with the criterion 'benefit to patient', the levels of 'small', 'medium' and 'large' are intrinsically ranked, with 'small' relating the lowest value and 'large' relating to the greatest value. However, with the criterion 'age' the ordering of the levels is not obvious. Some people may consider the elderly to have more value than the young for instance.

In this thesis the age levels in the decision survey are ranked in ascending order of importance from old to young – 65+ years, 15-64 years and 0-14 years[97] – reflecting the FIA. The three age groups broadly encompass three segments of society: retired people, working-aged people, and babies and children. However, future research could be done to determine which age groups New Zealanders value the most and then narrower age groups could be used in future decision surveys.[98]

Age of patient:
(levels in ascending order of importance)

- 65+ years
- 15-64 years
- 0-14 years

4.3.2 Alternative treatments

'Alternative treatments' encompasses two aspects. First, that another treatment is available, albeit a less effective one, and second, that people can afford to pay for the treatment if it were not publicly funded.

[97] The New Zealand Treasury used these three age groups in their report "Population Ageing and Government Health Expenditure" (New Zealand Treasury 2005).

[98] It would also be interesting to compare New Zealanders' preferences with respect to age with findings from international studies. For instance, some cultures value their elderly more than other cultures. Also, when life expectancy significantly varies between ethnicities, an age group that might be considered 'old' in one ethnicity may not be considered 'old' in another.

> "[with respect to PET scans[99]]... we already have quite good MRI and CT scans and so we have quite good alternatives. I think it's being seduced by technology just because it's new and fancy technology."
>
> *Member of the GP practice focus group*

> "I did think about the ones [treatments] where maybe people should possibly pay for them themselves. Whether that is more appropriate, rather than society looking after everyone, that the individual should look after the situation and pay for it themselves."
>
> *Member of the non-medical health workers' focus group*

It is difficult to assess public affordability of health care treatments. Even relatively cheap treatments may be out of reach to people on a low income. For this reason 'alternative treatments' in this thesis refers to whether another treatment is available.

Alternatives to currently available treatments are continuously becoming available due to advances in technology. New health technologies including vaccines, drugs, surgical procedures and equipment increase the treatment options available to patients but raise reimbursement and access issues. Patients become aware of new treatment options, predominantly by researching on the internet, and demand the best. However, developing and testing new technologies is a long and expensive process and the reimbursement required by companies is often prohibitive for health-funding organisations particularly when the technology is for a rare disease as the cost of treatment is spread over a small number of people.

Funding treatments for patients when no other alternatives exist, particularly when the opportunity cost of funding those treatments is high, is an issue New Zealand's PHARMAC (and other health-funding organisations) contends with. Based on cost-effectiveness alone most orphan drugs[100] would not be funded. Denis et al. (2009) suggests that additional criteria such as severity of disease and the availability of other treatments become more important in reimbursement decisions of orphan drugs, because of their comparative cost-ineffectiveness. McCabe (2005) questions whether it is fair that society place a higher value on the health of a

[99] PET Scans (Positron Emission Tomography) are a sensitive form of x-ray scanning used to detect diseases such as cancer, some heart disease and brain abnormalities in order for clinicians to plan the best form of treatment.

[100] Orphan drugs are drugs or medicines used to treat rare diseases.

person with a rare disease than the health of a person with a common disease. Nevertheless it is an issue that health-funding organisations need to address.

In New Zealand, from 1 March 2012, patients with rare disorders can apply to the Named Patient Pharmaceutical Assessment (NPPA) for unfunded drugs, a scheme operated by PHARMAC. This scheme replaces the Exceptional Circumstances Scheme.[101] Patients can assess funding through three channels: Unusual Clinical Circumstances, Urgent Assessment and Hospital Pharmaceuticals in the Community. The New Zealand Ministry of Health also sets money aside in a Special High Cost Treatment Pool for one-off treatments. Applications for assistance from the Special High Cost Treatment Pool are made through District Health Boards on behalf of patients, for treatments that are currently only available outside of New Zealand, or outside the public health system.

A report to the Minister of Health in New Zealand on access to high-cost, highly-specialised medicines (McCormack et al. 2010) makes 17 recommendations. Two of these recommendations are:

> "That prioritisation and funding decisions concerning high-cost, highly-specialised medicines continue to be made in the same way as such decisions for other medicines ... To be clear, we do not recommend that new and separate prioritisation processes and pools of funding be established for high-cost, highly-specialised medicines". (p 15)

> "That New Zealand-based public good research be applied to identifying targetable individuals with specific characteristics who are expected to receive the benefits from high cost medicines". (p 21)

If it is accepted that treatments for rare diseases are to be considered alongside treatments for more common conditions, the benefits from the treatments need to be separated from cost.

[101] Under the Exceptional Circumstances Scheme, approval for funding expensive medicines, including orphan drugs, could be obtained through one of three ways: the Community Exceptional Circumstances, the Hospital Exceptional Circumstances and the Cancer Exceptional Circumstances, where specific criteria had to be met before funding was considered. I spoke with a mother who has two sons with a lypsosomal disease. This disease is treated with enzyme replacement therapy, a very expensive treatment. To be eligible for funding under the Community Exceptional Circumstances scheme, there had to be fewer than 10 people in New Zealand with the disease. Though there were currently fewer than 10 people diagnosed with the disease in New Zealand, the two boys did not fall into the category of "exceptional circumstances" because statistically there should have been more than 10 people with the disease (i.e. some patients may be asymptomatic so have yet to be diagnosed) and therefore funding was not available. She believes it is a lottery as to who gets funding unless families have the support of a politician, a lobby group or a drug company. However, she believes that the allocation of funding for rare diseases should follow the same prioritisation process as that used for common diseases and that they should not be considered in a pool by themselves. It is clear that funding for treatment of rare diseases is an emotional and challenging area and one in which further work needs to be done.

This is consistent with the MCDA framework of this thesis where the benefits of treatment are considered before cost is introduced.

Treatment options for this patient:
(levels in ascending order of importance)

- this is the best treatment (there are less effective alternatives)
- this is the **ONLY** treatment available

4.3.3 Equality/equity/access

> "We are trained to treat everyone equally so we don't think about how or why they are seeking treatment."
>
> *Member of the nurses' focus group*

> "We have an obligation as a caring nation."
>
> *Member of the non-medical health workers' focus group*

Health inequality refers to differences in the health status and/or in the distribution of health resources between different population groups (WHO 2011); for example, people from different socioeconomic groups having different life expectancies. Some health inequality is unavoidable; for instance, it may be impossible or unacceptable to alter the distribution of health resources to reduce inequalities resulting from biological differences or from people making free choices that negatively impact their health. On the other hand, some health inequality results from factors that are avoidable and unnecessary such as uneven access to health and other publicly-provided services. This type of inequality can lead to inequity in health (Whitehead 1991).

Whitehead (1991) argues that equity does not mean that everyone should have the same health status or use the same amount of health resources irrespective of need. Inequity has a "moral and ethical dimension" and differences in inequality that are "unnecessary and avoidable" are therefore considered to be "unfair and unjust" (p 219). As a working definition, Whitehead defines health equity as: equal access to available care for equal need, equal utilisation of resources for equal need, and equal quality of care for all. No one should be disadvantaged from attaining their full health potential if it can be avoided. Whitehead emphasises that the problem of inequity needs to be tackled at an overall level, not just by one

sector, and that people need to be encouraged to "participate in every stage of the policy-making process" (p 223).

Culyer (2001) asserts that there is "no single universal theory of equity but it is widely agreed that equity implies equality", although there is no agreement on "*what* should be equal" (p 275). Culyer & Wagstaff (1993) explores four definitions of health care equity: equal utilisation, distribution according to need, equal access and equal health outcomes. Irrespective of how need and access are defined, the authors argue that the four definitions are incompatible. They contend that equality of health is the dominant principle underlying equity in health care and therefore the main focus should be on distributing health services in such a way to achieve "equal distribution of health" (p 431). With respect to the classical Aristotelian concepts of horizontal and vertical equity, individuals with equal needs should be treated the same (horizontal equity) whereas individuals with greater needs should receive more resources, in proportion to their greater need (vertical equity) (Culyer 2001).[102]

Braveman & Guskin (2003) argue that to achieve equity in health, all population groups need equal opportunities to be healthy. Resources across all policies and programmes, not just health, need to be distributed in such a way to equalise the health outcomes of the disadvantaged social groups with the more advantaged social groups.

Woodward & Kawachi (2000) advocate policies such as investment in human capital, redistributive policies and comprehensive access to health care as ways to achieve greater equity in health and as a result, reduce the spill-over effects of socioeconomic inequality.

Reducing inequalities is a priority for the New Zealand Government (Ministry of Health 2000). They recognise that there are significant inequalities in health based on ethnicity, socioeconomic factors, location and gender. In 2004 a series of workshops was held with staff from the health sector to increase the knowledge and skills of the participants with respect to inequalities in health. As a result, the Ministry of Health developed The Health Equity

[102] In the wider literature on social and distributive justice, John Rawls established principles relating to social and economic inequalities including 'the difference principle' which asserts that the least-advantaged members of society should receive the greatest benefit.

Assessment Tool (HEAT) for Tackling Inequalities in Health, which is to be considered alongside the Ministry of Health's Intervention Framework (Signal et al. 2007).[103]

In New Zealand there is a clear disparity in health between Maori and non-Maori.[104] This is discussed in the next sub-section.

4.3.3.1 Maori/non-Maori

> "Most things, if they were delivered appropriately so that Maori were benefiting from them, would benefit Maori more than everyone else because their health statistics are often poorer so all these sorts of things could benefit Maori and improve equity if they were properly delivered."
>
> *Member of the public health focus group*

> "The whole reason for having Whanau Ora is because Maori are not accessing the mainstream. Until things get better we need that."
>
> *Member of the Maori health provider focus group*

> "Culturally appropriate whatever that culture is – that's not just ethnicity."
>
> *Member of the Maori health provider focus group*

> "If they had an assessment programme that was holistic then they could determine that need for everybody regardless of race."
>
> *Member of the Maori health provider focus group*

Harris et al. (2006) found that inequalities in health outcomes between Maori and non-Maori are contributed to by a combination of deprivation and "experiences of perceived racial discrimination" (p 2007). Whether these inequalities are inequitable however, depends on whether the inequality results from factors that are avoidable or unnecessary and therefore are considered to be unfair or unjust.

---

[103] HEAT is a set of questions designed to assist health practitioners in understanding health inequalities when making decisions. A copy of the Health Equity Assessment Tool (Equity Lens) for Tackling Inequalities in Health is in Appendix 4.1

[104] For example, in 2006 the life expectancy for Maori at age 50 was at least six years less than that for non-Maori for both genders; lung cancer mortality was six times higher for Maori females aged 50-64 years than for non-Maori females of the same age group; the rate of renal failure with concurrent diabetes was 12 times higher in Maori males aged 50-64 than non-Maori males of the same age group (Ministry of Health (2011).

The United Nations Declaration on the Rights of Indigenous Peoples (United Nations 2007) and the Treaty of Waitangi give Maori the right to "monitor the Crown and to evaluation Crown action and inaction" (Robson & Harris 2007). In 1988 a set of principles was identified by the Royal Commission on Social Policy to clarify the relationship between the Treaty and health (Kingi 2007). These principles are 'partnership' (including Maori in the decisions and plans relating to Maori health), 'protection' (ensuring that the outcomes for Maori and non-Maori are the same) and 'participation' (encouraging Maori involvement in the planning, design and delivery of health services).

A Maori health model, whare tapa wha, compares health to the four walls of a house, each side representing a different dimension of health: spiritual, thoughts and feelings, physical and family. In a study by Devlin et al. (2000) participants were asked to score their own health and 14 core EQ-5D (or EuroQoL Group) health states on a visual analogue scale[105] and to comment on whether the questionnaire covered all aspects of health that were important to them. In a follow-up study by Perkins et al. (2004), the authors found that the concept of spiritual health was raised as frequently by non-Maori as Maori and concluded that most Maori do not conform to the Maori health model.

The Maori health provider focus group in this thesis, acknowledged that the health needs of Maori are often greater than those of non-Maori and that the way health services are delivered contributes to the inequality. However, the group did not think that Maori should be given preference based on ethnicity. The group advocated a collaborative approach between primary and secondary services. They also suggested that health services should be delivered in a way that encourages Maori to utilise them; for example, providing free transport to and from health centres, sending appointment reminders and encouraging patients to bring whanau to appointments.

Health inequalities will always exist but they may not necessarily be inequitable – the distinction is whether the inequality is caused by factors that are unnecessary or avoidable. In these circumstances equity can be enhanced by ensuring effective health care delivery and addressing broader issues such as employment, education, social and cultural wellbeing, for both Maori and non-Maori. As such, equality/equity/access will not be considered as a separate criterion in the decision survey. However, aspects of equality and equity are

[105] A visual analogue scale (VAS) is an instrument used to measure a characteristic or attitude across a continuum of values; in this case, particular health states.

encompassed in the criteria 'need' (greater consideration is given to people with comparatively greater health needs) and 'lifestyle choice' (lower consideration is given to people who have become ill through the 'avoidable' choices they have made), discussed in Sections 4.3.4 and 4.3.8 respectively.

4.3.4 Severity-of-illness/need/treating the sickest first

> "It's the law of the jungle, the strongest survive – that's healthy for society."
>
> *Member of the retirees' focus group*

> "Preventing death is more important than reducing an illness which is more important than enhancing your life..."
>
> *Member of the public health focus group*

According to Cookson & Dolan (2000), rationing decisions require the *relative* degree of need to be considered as well as the need itself. The authors found that 'need' defined in terms of ill health, is broadly understood but that 'need' should be interpreted to include the capacity of an individual to benefit from treatment. Without this consideration, treating some individuals in great need may be futile.

When developing standardised measures for prioritising patients on waiting lists, Hadorn et al. (2000) considered 'need' to be equivalent to a combination of 'severity and urgency' and that other non-clinical factors such as the patient's ability to work or to live independently could be included in defining the level of 'need'. The key difference between 'severity and urgency' is the expected benefit of treatment. For example, early stage cancer might not be considered 'severe' but if left untreated the benefit of treatment would be reduced and therefore it could be considered 'urgent'.

'Severity-of-illness' is widely regarded as a relevant criterion in prioritising patient groups (Nord et al. 1996). The severity-of-illness approach takes into account not only the severity at the time of intervention but also *expected* severity in years to come (Nord 2005) and suggests that priority should be given to patients who have the poorest health prospects, particularly when their ability to function in daily life is severely inhibited (Stolk et al. 2005).

On the other hand, the 'rule of rescue' equates to giving preference to the most severe health conditions (James et al. 2005). The 'rule of rescue' refers to the ethical duty of society to do as much as possible for identifiable individuals in imminent danger of death, regardless of the costs involved or the size of the health gain (Hauck et al. 2004, Cookson et al. 2008).

The crucial distinction between the severity-of-illness approach and the 'rule of rescue' is that the severity-of-illness approach gives equal priority to individuals currently suffering from a severe illness with individuals who may become ill in the future whereas the 'rule of rescue' favours identifiable individuals in immediate danger of death over unidentified individuals who may become gravely ill in the future (Cookson et al. 2005).

Shah (2009) reviewed studies using 'severity-of-illness' as a factor in economic evaluations. The author found that, in general, people are willing to forfeit the goal of maximising societal health in order to give priority to the severely ill and that 'severity-of-illness' as a priority-setting criterion is widely supported.

In a report by the NICE Citizens' Council in 2006, it was suggested by some Council members that improving the quality of life of a dying person was just as important as saving their life in the first place.

The criterion 'patient's health before treatment' in this thesis encompasses the concepts of need, severity of illness, treating the sickest first, rule of rescue and palliative care.

Patient's health before treatment:
(levels in ascending order of importance)

- relatively good (though treatment is still beneficial)
- fair (neither good nor bad)
- poor (but not immediately life threatening)
- will die without treatment

4.3.5 Impact on family and/or society

> "The HPV one [cervical cancer vaccine] is not only for the women who have the vaccine but also for their possible 15 sexual partners for the rest of their lives so it's like a trickle-down effect for society."
>
> *Member of the Maori health provider focus group*

> "It [the treatment] was very important because it reached a large number of people, with flow-on positive effects of limiting illness."
>
> *Member of the GP practice focus group*

> "I ranked that very highly [postnatal depression] in terms of thinking not only about the benefit for the individual but what the impact might be – the impact on the child and the family and in the future and what impact that might have on society."
>
> *Member of the public health focus group*

In addition to the direct benefits an individual receives from health care, there may be flow-on effects (externalities) that impact society. For example, methadone maintenance therapy for drug users who inject is associated with a substantial reduction in the costs of crime and imprisonment (Sheerin et al. 2004); a depressed mother receiving postnatal counselling receives a direct benefit from treatment but, in addition, her happier health state is beneficial to her baby and family.[106]

James et al. (2005, p 37) refers to the "external impacts or spill-over of a disease" when exploring factors that could be included in priority-setting and suggests that external impacts should be incorporated when measuring the cost-effectiveness of treatment.

Cookson & Dolan (2000) discusses the "indirect consequences" of prioritising patients (p 326). When considering an experimental drug treatment for an 11-year old boy with cancer, the authors identify three indirect consequences of treatment: the knowledge gained from using experimental drugs may be of benefit to future patients; if treated successfully the boy might be able to have children in the future; and the boy's parents may suffer ill health if their child was denied treatment.

Boulier et al. (2007) attempted to quantify the magnitude of positive externalities associated with influenza and mumps vaccinations in the US. The authors found that the size of the externality varied depending on the infectiousness of the disease and the effectiveness of the vaccine. However, in some simulations, for every vaccination given, more than one case of illness was prevented amongst unvaccinated people.

---

[106] Martire et al (2004) explored the link between chronic illness and family relationships. The authors found that family members who have to look after a sick family member often suffer poor psychological and physical well-being as a result.

Some of the flow-on effects mentioned in the focus groups were: previously sick people returning to work which increases the Government's tax intake, patients no longer requiring a carer, lower utilisation of health resources, preventing illnesses from spreading, reducing crime, and strengthening the whanau *(family)*.[107]

There is potential for the criterion 'benefit to others (*eg* family or society)' to be interpreted differently by respondents. However, I found that overall, participants in the focus groups and respondents in the pilot study[108] had a good understanding of 'benefit to others' and consistently interpreted it in a similar way.

Benefit to others (*eg* family or society):
(levels in ascending order of importance)
- small
- large

4.3.6 Quality of life/length of life

> "At my age if I had a choice between 5 and 10 [extra years] I would take 10. Every day above the ground is good for me and my family. Anything over 60 is good. Most of mine [whanau] are dead between 46 and 60."
>
> *Member of the Maori health provider focus group*

> "If they [the patients] are going to have quality of life for a number of years, it's a horrible thing to say but surely it's more important than, say, if somebody has only got a short time."
>
> *Member of the retirees' focus group*

> "We are taking away the choice of how they [the elderly] might wish to die."
>
> *Member of the non-medical health workers' focus group*

Some respondents commented that quality of life encompasses much more than just physical health and a holistic approach is required to measure HRQoL. Other respondents commented that although more advanced treatments are now available, life should not be prolonged at any cost.

---

[107] Another externality, not mentioned in the focus groups, is psychic externalities: people care about other people and benefit from seeing them treated.
[108] The criteria and levels were tested in a pre-test and a pilot study. This is discussed in the next chapter.

Respondents agreed that extending life and saving lives were important, but that these factors should not be considered in isolation from quality of life.[109] A treatment may extend life but if the quality of those additional life years is poor, then the benefit from treatment may be relatively small. On the other hand, improving the quality of life of a dying person may be of great benefit. Arguably, most people would prefer more life years than less, and a higher quality of life than a lower quality of life. Combining 'length of life' and 'quality of life' takes into account the quality of any additional life years gained from treatment.

As discussed in Chapter 2, a common way to combine the length and quality of life into one measure is to use QALYs which can then be used as a 'common currency' by which to compare treatments.

Two versions of the decision survey are developed: one for the general public and another for health services researchers. In the health services researchers' survey the criterion relating to individual benefit is labelled 'QALYs gained by treatment' and the three levels are: small (< 1 QALY), medium (1-3 QALYs) and large (> 3 QALYs). In the general public's survey the criterion is labelled 'benefit to patient' with three levels: small, medium and large. This is because most members of the general public do not understand what a QALY means without detailed explanation. The criterion is therefore simplified to make it easier for the respondents to understand. There is a risk that respondents might interpret these levels differently in terms of length and quality of life. However, respondents in the pre-test and pilot surveys consistently interpreted the levels in the same way. This is discussed in the next chapter.

Benefit to patient (*ie* length and/or quality of life):
(levels in ascending order of importance)

- small
- medium
- large

[109] The quality of life of people with disabilities was never raised in the focus group meetings, possibly because there were no vignettes that were associated with permanent disabilities. To ensure that the trade-off survey encompassed all possible considerations, I interviewed Donna-Rose McKay, the head of Disability Information and Support at the University of Otago, to obtain her views on whether the health needs of disabled people should be a separate consideration within a prioritisation framework. Ms McKay believes that people with disabilities should be treated the same as everyone else. However, as disabled people often have minor ailments that can very quickly lead to serious illness, their health needs may be comparatively greater than able-bodied people. She suggested that is the *delivery* of health services that is important – disabilities need to be identified and correctly diagnosed to ensure that disabled people receive appropriate treatment to remain healthy thereby reducing the associated costs of lost work and additional treatment.

4.3.7 Preventive health care

> "It's a burden on your infrastructure – you have people waiting for 10 years until they can have their hip replacement – going to doctors, taking lots of pain killers, clogging up health systems, with treatable things. It flows on to their mental health as well – I've had one old lady say to me you wouldn't treat a dog like this – why won't someone take me out and shoot me because she was still on the waiting list."
>
> *Member of the Maori health provider focus group*

Promoting preventive health measures and early intervention were two issues frequently raised in the focus groups. According to Russell (2007), preventive health measures include "vaccines that prevent disease, medications that reduce the risk of developing disease, screening tests that detect diseases at an early stage when treatment is more effective, and lifestyle changes – smoking cessation, exercise, diet – that keep people healthy longer" (p 1).

To evaluate the cost-effectiveness of preventive health measures, the cost of the intervention is compared with the savings in medical costs as a result of the intervention. To be most effective, however, the *actual* people who will require health treatment in the future need to be identified, which is difficult. People *at risk* of future health problems can be identified; for example, obese people have a greater risk of developing diabetes and other health-related problems. However, the cost of screening and providing medication to the entire group of 'at risk' people in order to identify those people who will become ill may far outweigh the benefits (Russell 2007). Sometimes it is more cost-effective to treat people as they become ill instead of providing preventive health care (Cohen et al. 2008).

The difference between preventive treatments and non-preventive treatments (such as treatments that cure or prevent disease progression, for example) is the timing of the benefits. Preventive treatments accrue benefits in the future in terms of HRQoL and/or length of life to the patient and in some cases positive flow-on effects to society, whereas non-preventive treatments yield benefits almost instantaneously. Therefore preventive health care is not a separate criterion but is implicitly included as part of 'benefit to patient' and 'benefit to others'.

4.3.8 Effectiveness

> "What is the relative success of the treatment − comparing treatments that cure with those that extend life, etc − and then separately [consider] the quality of medical evidence."
>
> *Member of the non-medical health workers' focus group*

> "I think it's important to distinguish between possible or likely efficacy and effectiveness and proven efficacy and effectiveness."
>
> *Member of the public health focus group*

When discussing the effectiveness of health treatments, focus group participants talked about the success of treatment, duration of treatment, side effects, short-term versus long-term effectiveness and cure versus maintenance. The effectiveness of a treatment is related to two criteria: 'benefit to the patient' (the more effective a treatment is, the greater the benefit to the patient) and 'treatment options for this patient' (sometimes there are alternative but less effective treatments available).

'Quality/strength of medical evidence' relates to the safety and efficacy of health treatments and is an important consideration in the priority-setting process. 'Quality/strength of medical evidence' is not a separate criterion in the decision survey but will be considered alongside cost and other additional factors. This is explained in Chapter 9.

4.3.9 Lifestyle choice

> "Is there a standard lifestyle out there and what is it? What's normal? Like $80K, dog Chucky and cat Bollie? You can't do that. That's why we work where we do because we know that's not the case. We have single 17-year old parents, some fantastic and some that aren't. You can't generalise. Everyone has different resources, different backgrounds, different strengths and weaknesses."
>
> *Member of the Maori health provider focus group*

> "Most of the people who need it [health care] created the problem themselves."
>
> *Member of the retirees' focus group*

> "Statins – you are giving them a pill to lower cholesterol. It's taken care of, they are not responsible anymore so they don't change."
>
> *Member of the non-medical health workers' focus group*

The impact a person's lifestyle has on their health status was raised as an issue by all of the focus groups bar the public health group. Four of the five groups felt that people are responsible for their own health and that preference should be given to individuals whose lifestyle choices have *not* contributed to their illness or injury. The Maori health provider group, while acknowledging that a person's lifestyle contributes to their health status, felt that not everyone has the same opportunities (e.g. limited resources) and therefore people who become ill through lifestyle 'choices' such as smoking or over-eating, should not be penalised.

There is some support in the literature for prioritising patients based on their lifestyle choices. For example, Nord et al. (1995) surveyed a cross-section of Australians regarding their attitudes to health maximisation and egalitarianism. In response to a question related to lifestyle, many of the respondents felt that some priority should be shown to non-smokers over smokers. The authors found that of the 59.5% of respondents who wanted to give priority to non-smokers, 31% were smokers themselves. Bowling (1996) elicited the views of a large sample of the general population in the UK about priorities for health services. The author found public support for giving lower prioritisation to people who have 'self-inflicted' conditions, for example, smoking.

James et al. (2005) refers to "collective versus individual responsibility" (p 43). Priority is often given to health treatments that favour the most disadvantaged members of society but the authors question whether society should contribute to the cost of treatment for individuals who through their 'voluntary' behaviours are directly responsible for their lower health status.

When defining equity in health care, Whitehead (1991) differentiates between what is "inevitable or unavoidable" and what is "unnecessary or unacceptable" (p 219). Illness resulting from a severely restricted lifestyle – for example, a lack of resources, little or no social support, inadequate housing, unemployment, limited access to health and other public services – is unnecessary and therefore unfair. On the other hand, health-damaging behaviour which is 'freely chosen' such as smoking or participating in risky sports is avoidable and therefore any illness that results from the behaviour is *not* unfair.

Woodward & Kawachi (2000) distinguish between inequality in health that has arisen from circumstances where people have little or no responsibility for their actions and where people are fully responsible for the actions leading to a loss in health status. The authors suggest that health inequalities can be reduced if resources are directed to those people who are suffering ill health as a result of circumstances in which they have little control. However, it is difficult to define the extent to which people can be held responsible for their actions. They use tobacco smoking as an example. Whether a person freely decides to start smoking is a consequence of both the environment and their personal choice. For instance, advertising, availability, role models and adolescent peer pressure contribute to the decision to start smoking, and because of its addictive nature, the limits of 'personal choice' are questionable.

Jusot et al. (2010) identify two aspects of health inequalities in France: "inequalities of opportunities" and "inequalities due to differences in effort" (p 1). "Inequalities in opportunities" refer to an individual's personal circumstances such as their family background, whereas "inequalities due to differences in effort" refer to factors for which an individual is responsible, such as exercising and not smoking. Although these two aspects were difficult to separate Jusot et al. attempted to quantify the "inequality of opportunity" component of health inequality. The authors found that approximately 45% of health inequality is due to "inequality of opportunity". Jusot et al. conclude that a parent's social background[110] is a significant determinant of their children's health and emphasise the intergenerational transmission of lifestyles particularly in relation to smoking and diet.

At a policy level it might be difficult (and contentious) to identify individuals who 'voluntarily' contribute to their illness. However 'lifestyle choice' was raised as an important issue in most of the focus groups and is included as a criterion in the decision survey.

Illness or injury caused mainly by lifestyle choices:
(levels in ascending order of importance)

- yes
- no

[110] A parent's social background is determined by their education, work status, health, financial circumstances growing up and whether they smoke and/or drink alcohol.

4.3.10 Number of people

> "The number issues ... you balance the number of people it's going to affect with whether it is going to save their life or just make them feel slightly better and the size of the effect."
>
> *Member of the public health focus group*

> "It was the numbers that I sort of looked at sometimes. That's a lot of people to benefit from that."
>
> *Member of the retirees' focus group*

> "Because of the numbers – there were big numbers – potentially preventing 66,000 heart attacks."
>
> *Member of the GP practice focus group*

As discussed in Chapter 3, participants in the focus groups were asked to rank 14 vignettes describing a variety of health treatments before attending their focus group meeting. Included in each vignette was the number of patients who would receive treatment over a one-year period and the benefit to those patients of receiving treatment. It was evident by the comments made in the focus group meetings, that the number of patients to be treated greatly influenced the rankings. For example, one vignette, "statins for patients at high risk of cardiovascular disease" stated that if 220,000 patients were to receive statins for the rest of their lives, potentially 66,000 heart attacks or coronary deaths could be prevented. This vignette was ranked first by four groups and second by two groups.

In preference studies involving choices on two dimensions, Slovic (1975) and Tversky et al. (1988) found that when faced with two equally attractive options, individuals choose the option that is 'superior' on the most important dimension to the individual. This is known as the 'prominence effect'.

However, when decisions are complex participants may favour one criterion, not because of a particular preference, but in order to simplify and speed up the decision process. Favouring the treatments which affect the greatest number of people is referred to by Bryan & Roberts (2008) as the "numbers game nature of a discrete choice approach" (p 150).

Including 'number of patients' as a criterion in the decision survey might encourage some respondents to ignore the other criteria and always choose the option with the highest number of patients, as occurred when ranking the vignettes. To make sure that this does not happen, the number of patients affected by a treatment will be considered separately along with cost (discussed in the next section). This means that in the decision survey, respondents will be presented with hypothetical choices which involve two imaginary *patients* (who differ in only two characteristics) and not two *treatments* that involve a different number of patients. Including the total number of patients as a separate consideration is discussed in Chapter 9.

4.4 Cost

> "I found it very difficult to disregard cost and surely when you are planning, cost is a huge part of it."
>
> *Member of the GP practice focus group*

When the focus group participants were asked to rank the health vignettes, they were instructed to ignore cost and to consider only the treatment's benefits or value to society. The main reason for this, apart from cost being an obvious criterion, was to make it cognitively easier for the participants to rank the vignettes. As each vignette differs in terms of the number of patients being treated, including overall treatment costs would have made the ranking decision more difficult for some participants. For example, participants would need to compare the cost effectiveness of 7000 hip replacements costing $119,000,000, with dialysis for renal disease for 440 patients costing $22,000,000 with abatacept for rheumatoid arthritis for 30 patients costing $900,000.

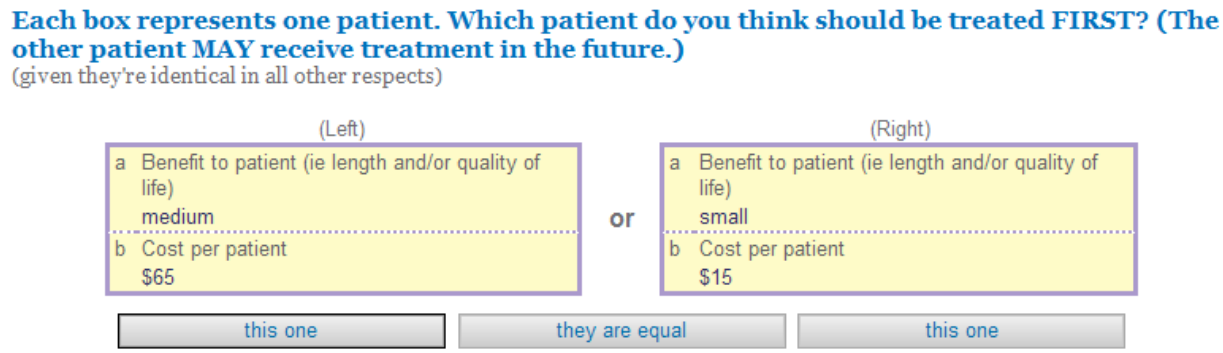
Similarly, if the total cost of treatment (i.e. the cost of the entire treatment programme or service) is included in the decision survey as one of the criteria, then arguably the total number of patients should also be included. For example, consider two treatment programmes, A and B. Programme A generates a large benefit and its total cost is $119,000,000; Programme B generates a medium benefit and its total cost is $900,000. A respondent could choose their 'preferred' programme without knowing how many patients are being treated simply based on the benefit and the cost. However, if the number of patients being treated differs between programmes then this could influence a respondent's decision (e.g. Programme A treats 1,000 patients and Programme B treats 10,000 patients). But adding

the total number of patients to each choice increases the complexity of the decision respondents are required to make.

One way to avoid this issue is to use 'cost per *patient*' instead of 'cost per *treatment programme*'.[111] However, using 'cost per patient' also creates difficulties. Not all treatments can be purchased (and provided) on an individual basis. For example, the cost of a cervical cancer vaccine is $350. However, this is based on 50,000 girls being vaccinated over five years at a total cost of $17,500,000. (Arguably the vaccine could be offered to fewer than 50,000 girls but the cost is likely to be greater than $350.) In contrast, a hip replacement costs $17,000 regardless of whether 10 or 1000 are performed. Therefore it is impractical to use 'cost per patient' without including the number of patients. Cost per patient alone, provides little insight into the economic impact of a treatment or whether it provides value for money (Menon & Stafinski 2008).

Including cost as a criterion in the decision survey not only adds complexity to the trade-off questions but also creates uncertainty relating to opportunity cost (Bryan et al. 2002). For example, consider the two options in Figure 4.1.

Figure 4.1: Trade-off question including cost



The treatment for the patient on the left costs $65 and the treatment for the patient on the right costs $15. If the patient on the right is treated (i.e. $15) does this mean that $50 has been saved or does it mean that more of that treatment can be purchased or does it mean that the treatment for the patient on the left (i.e. $65) is more valuable than that the treatment on the right (i.e. $15)?

[111] As discussed in the previous section, the trade-off survey will include choices involving two *patients* and not two treatment programmes. Therefore cost per patient and not cost per treatment is appropriate.

Cost (and other considerations such as the number of patients and strength of medical evidence) will be included as separate considerations within the prioritisation framework. This will be explained in Chapter 9.

4.5 Criteria and levels for the decision survey

In the previous sections the suggested criteria from the focus groups, criteria from the literature and comments from health experts were discussed. From eleven broad categories, six main criteria were established for the decision survey. These criteria and their associated levels are listed below in Table 4.3.

Table 4.3: Criteria and levels for the decision survey[112]

Treatment options for this patient
this is the best treatment (there are less effective alternatives)
this is the **ONLY** treatment available
Age of patient
65+ years
15-64 years
0-14 years
Benefit to others (*eg* family or society)
small
large
Patient's health before treatment
relatively good (though treatment is still beneficial)
fair (neither good nor bad)
poor (but not immediately life threatening)
will die soon without treatment
Benefit to patient (*ie* length and/or quality of life)
small
medium
large
Illness or injury caused mainly by lifestyle choices
yes
no

[112] The levels are in ascending order of importance.

It is important that the criteria and levels for the decision survey are concise and easy to understand to reduce respondent burden when answering the questions. It is also important that respondents interpret the criteria and levels in a similar way.[113]

As mentioned in Sections 4.3.1 and 4.3.8, it may be difficult in practice to use the criteria 'age of patient' and 'illness or injury caused mainly by lifestyle choices' because of discrimination or in the case of lifestyle, for equity reasons. However, as these issues were raised in every focus group, it will be interesting to discover how important these criteria are to the general public.

4.6 Conclusion

In this chapter the process of establishing the relevant criteria and associated levels to include in the decision survey was explained. Potential criteria elicited from studies relating to health care prioritisation were amalgamated with the feedback from the focus groups and expert opinions to form six criteria. The implementation of the decision survey is explained in the next chapter.

> "I'm glad I don't have to make these decisions."
>
> *Member of the retirees' focus group*

[113] A pre-test and pilot study were done to ensure that the criteria and levels were easily understood, that the format and instructions of the survey were easy to follow and that the criteria were being interpreted in a similar way by the respondents. This is discussed in the next chapter.

Appendix 4.1: HEAT Tool

A Health Equity Assessment Tool (Equity Lens) for Tackling Inequalities in Health (May 2004)

There is considerable evidence, both internationally and in New Zealand, of significant inequalities in health between socioeconomic groups, ethnic groups, people living in different geographical regions and males and females (Acheson 1998; Howden-Chapman and Tobias 2000). Research indicates that the poorer you are, the worse your health. In some countries with a colonial history, indigenous people have poorer health than others. Reducing inequalities is a priority for government. The New Zealand Health Strategy acknowledges the need to address health inequalities as 'a major priority requiring ongoing commitment across the sector' (Minister of Health 2000).

Inequalities in health are unfair and unjust. They are also not natural; they are the result of social and economic policy and practices. Therefore, inequalities in health are avoidable (Woodward and Kawachi 2000).

The following set of questions has been developed to assist you to consider how particular inequalities in health have come about, and where the effective intervention points are to tackle them. They should be used in conjunction with the Ministry of Health's Intervention Framework (Ministry of Health 2002).

1. What health issue is the policy/programme trying to address?
2. What inequalities exist in this health area?
3. Who is most advantaged and how?
4. How did the inequality occur? (What are the mechanisms by which this inequality was created, is maintained or increased?)
5. What are the determinants of this inequality?
6. How will you address the Treaty of Waitangi in the context of the New Zealand Public Health and Disability Act 2000?
7. Where/how will you intervene to tackle this issue? Use the Ministry of Health Intervention Framework to guide your thinking.
8. How could this intervention affect health inequalities?
9. Who will benefit most?
10. What might the unintended consequences be?
11. What will you do to make sure it does reduce/eliminate inequalities?
12. How will you know if inequalities have been reduced/eliminated?

(Adapted from Bro Taf Authority. 2000. *Planning for Positive Impact: Health inequalities impact assessment tool*. Cardiff: Bro Taf Authority.)

Amended by Ministry of Health. May 2004.

Source: Te Roopu Rangahau a Erü Pomare., Ministry of Health and Public Health Consultancy. 2003. *A Health Equity Assessment Tool*. Wellington: Public Health Consultancy, Wellington School of Medicine and Health Sciences.

References

Acheson D. 1998. *Independent Inquiry into Inequalities in Health*. London: The Stationery Office.

Howden-Chapman P, Tobias M (eds). 2000. *Social Inequalities in Health: New Zealand 1999*. Wellington: Ministry of Health.

Minister of Health. 2000. *The New Zealand Health Strategy*. Wellington: Ministry of Health.

Ministry of Health, Public Health Consultancy, Te Roopu Rangahau Hauora a Eru Pomare. 2002. A Health Equity Assessment Tool. Department of Public Health, University Otago and Ministry of Health

Woodward A, Kawachi I. 2000. Why reduce health inequalities? *Journal of Epidemiology and Community Health* 54(12): 923.

~ Chapter 5 ~

The decision survey and sample groups

5.1 Introduction

In this chapter the development of the online survey is explained and the process of establishing the sample groups is described. Issues with sample selection and potential problems with the survey are covered later in the chapter. The chapter concludes with a discussion on the validity and reliability of the survey including the results of a 'test re-test'.

5.2 Decision survey

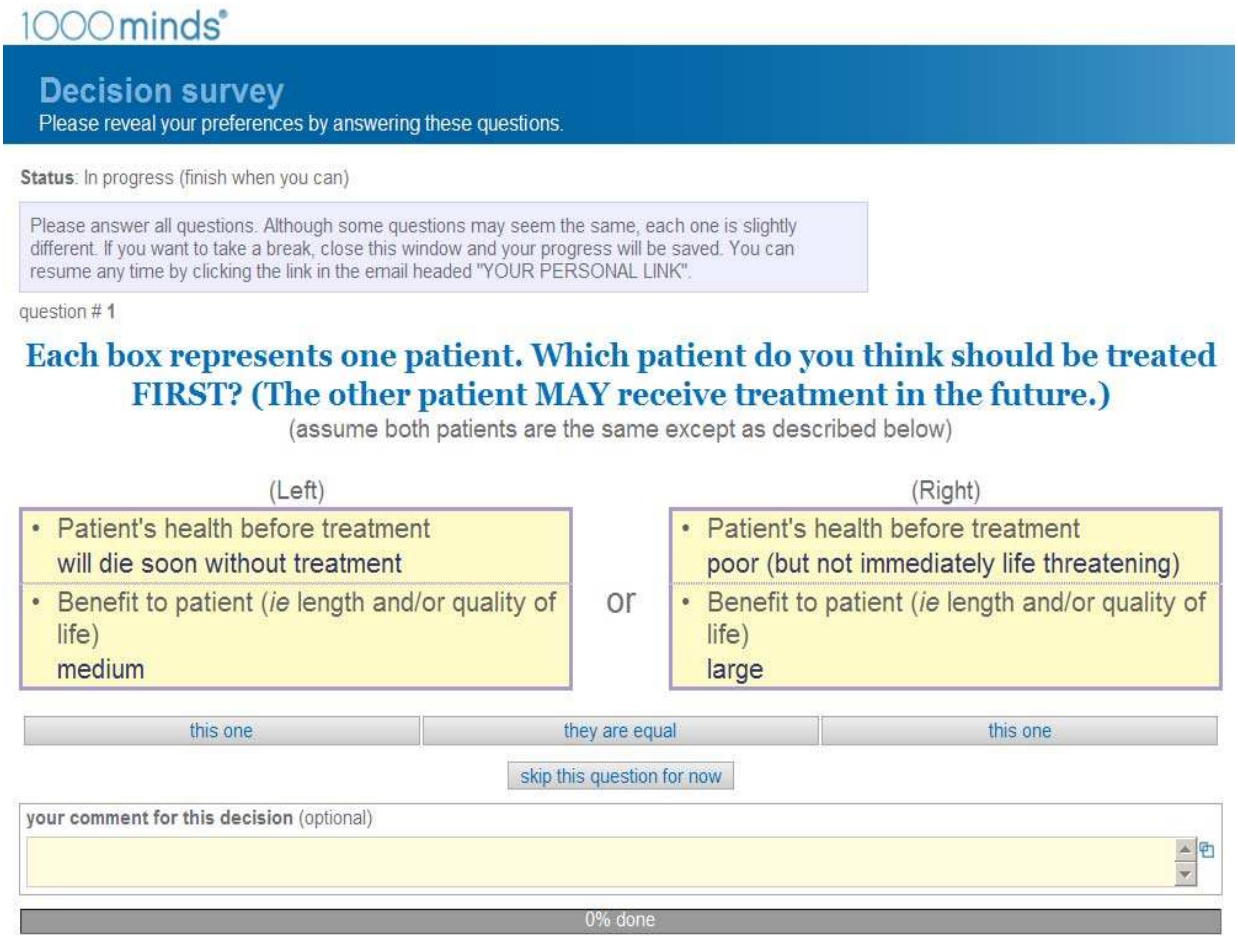
The primary objective of the online decision survey is to elicit the preferences of each of the respondents with respect to the relative importance of the six criteria discussed in the previous chapter. To simplify discussions relating to the criteria, the criteria from the decision survey are abbreviated as shown in Table 5.1.

Table 5.1: 'Abbreviated' criteria

Criteria from the decision survey	'Abbreviated' criteria
Patient's health before treatment (health status)	Need
Benefit to patient (*ie* length and/or quality of life)	Individual benefit
Age of patient	Age
Benefit to others (*eg* family or society)	Societal benefit
Illness or injury NOT caused by lifestyle choices	Lifestyle
Only treatment option available for this patient	Only available treatment

The decision survey comprises two parts. The first part of the survey consists of a series of hypothetical choices. Respondents are asked to trade-off one criterion with another, by choosing which of two imaginary patients should receive priority for treatment. An example of a trade-off question is in Figure 5.1. (Figure 5.1 is a screenshot of the first 'page' of the survey. When a respondent clicks on the link to the decision survey, this is what they see.)

Figure 5.1: Example of a question from the decision survey



In the second part of the survey respondents are asked questions relating to their demographic characteristics such as their age, income, education, health status, and so on. (Questions relating to sex, age, ethnicity, qualifications, income and household composition follow the 2006 Census statistical standards.) A copy of the demographic part of the survey is in Appendix 5.1.

Google Documents (https://docs.google.com) was used to create the demographic part of the survey. The survey link from the Google questionnaire is embedded in the 1000Minds survey. From a respondent's point of view the two sections are seamless and appear to be one survey.

Below each trade-off question is a space for respondents to comment on the decision they just made, should they wish to do so. At the end of the survey, respondents are invited to comment on the survey overall and/or on health prioritisation in general.

5.3 Pre-test and pilot study

Before undertaking a pilot study, the survey was pre-tested on10 respondents. Their feedback was sought with respect to the language used, the understandability of the criteria and levels, and the format and flow of the survey itself. In addition, I sat alongside five of the respondents while they completed the survey and had them verbalise their decisions. I did this to ensure that the respondents understood the choices they were being required to make and that they were interpreting the criteria in a similar way. For example, the criterion 'benefit to patient' has three levels: 'small', 'medium' and 'large'. These levels could be expanded to include both HRQoL and length of life aspects. However, it would make each trade-off question longer and more cumbersome to read which would likely lead to respondent fatigue. When respondents in the pre-test were asked to interpret the levels, it was apparent that they all had similar understandings of 'small', 'medium' and 'large' in terms of HRQoL and length of life.

A few minor changes were made to the wording of the criteria as a result of the feedback and a pilot study was undertaken.

Twenty two respondents consisting of colleagues, friends and family, completed the pilot survey. Some respondents wrote comments under the trade-off questions and/or at the end of the survey. All comments were followed up by an informal discussion with each respondent.

Most of the comments related to the wording of the criteria and the associated levels in the trade-off part of the survey, resulting in some minor changes being made. Some comments related to the content. For example, several people suggested that lifestyle should not be included in health prioritisation as it may disadvantage people who have not had the same life advantages as others. This issue was raised many times throughout the thesis and will be discussed further in Chapter 6.

Several other concerns were identified in the pilot study such as respondents disregarding the instruction to 'assume both patients are the same except as described below' and interpreting the criteria in different ways. These issues are discussed in Section 5.7.1.

5.4 Sampling

Various factors need to be taken into account when deciding which method of sampling to use; for instance, the type of survey (e.g. telephone, face-to-face, online),[114] cost, intended size of the sample and demographic representativeness.

In this thesis, obtaining a demographically representative sample is important for a number of reasons. Because health prioritisation, in essence, affects all New Zealanders in one way or another, it is important to survey the population as widely as possible to elicit the preferences of a broad cross-section of the New Zealand general public. When a sample is demographically representative the results of a survey can be generalised more reliably to the larger population.[115] It also allows the preferences of sub-groups within the population to be compared. For example, do younger people place more importance on the criterion 'age' compared to older people?; do Maori respondents place greater importance on the criterion 'need' compared to non-Maori respondents?

Two sampling approaches are used in this thesis: snowball sampling and random sampling. Snowball sampling was used to reach as many people as possible who might be interested in completing the survey. However, as snowball sampling can result in a non-representative sample (as discussed below), random sampling was also used to ensure that *potentially* everyone (over the age of 18) had an equal chance of being selected to complete the survey.

5.4.1 Snowball sample

Snowball sampling, also known as chain referral sampling, is often used in qualitative research where the focus of the study is particularly sensitive thus making it difficult to locate people for the study via more conventional sampling approaches (Biernacki 1981). Although health prioritisation could be considered to be a fairly contentious issue, it is not an overly sensitive issue. Nonetheless snowball sampling has a number of advantages. People interested in the survey can forward the survey link to other people interested in the survey, creating a fairly large sample group quickly and easily. A disadvantage of using snowball sampling in this area of research is that, because the sample is largely self-selected, the resulting sample is

[114] Survey design is discussed in Section 5.7.
[115] Other factors such as the response rate, affect whether the results of a survey can be reliably generalised to the wider population. These are discussed in Section 5.6.

most likely to be non-representative of the overall population as those more interested in health are more likely to have been contacted.

In this thesis, to 'start the snowball rolling', emails were sent to an initial group of respondents (i.e. friends and colleagues) inviting them to do the survey and to forward the email to anyone who they thought might be interested. An email was also sent to the departmental secretaries at the University of Otago with a request that they forward the email to their staff. Emails were also sent to focus group participants, the health specialists who commented on the vignettes and several other people who had indicated an interest in the research. The link was publicised in a primary school newsletter, and two newspaper articles relating to the survey were published in the *Gore Ensign* (a community newspaper) and the *Otago Daily Times*. Around 10 interested readers asked for a link to the survey. A copy of the email inviting people to do the survey is in Appendix 5.2 and copies of the two newspaper articles are in Appendices 5.3 and 5.4.

When potential respondents clicked on the link in the email, they were taken to the survey page and asked for their email address. The software automatically sent them an email with their 'personal link' to the survey.[116] If respondents wanted to take a break while completing the survey, they could close the survey and resume it later by clicking on their personal link.

Over a period of four weeks, 332 respondents obtained a personal link to the survey. As can be seen in Table 5.2, of the 322 respondents, 25 respondents did not start the survey, 32 respondents started the survey but did not complete it and 275 respondents completed the survey.

Table 5.2: Snowball sample respondents

Snowball Sample		
Total requests for survey	332	
Survey sent but not started	25	
Survey started but not completed	25	
Demographic part of survey not completed	7	
Total number of completed surveys		275

[116] A copy of the email containing a respondent's personal link is in Appendix 5.5.

5.4.2 Random sample

The main objective of undertaking the decision survey is to obtain the views of a representative sample of New Zealanders. Although the snowball sampling was convenient, time-saving and cost-effective, it did not produce a demographically representative sample.[117] To obtain a more demographically representative sample, thereby allowing the survey results to be generalised across the population overall, a wider cross-section of people needed to be surveyed. Therefore a stratified random sampling approach was taken.

Simple random sampling is characterised by potentially every person from a population having an equal chance of being selected. Although face-to-face samples often produce a reasonably demographically representative sample, online samples are generally considered biased, especially in terms of age, gender and education, because not everyone has access to a computer and/or the internet or has the confidence to complete an online survey (Blasius 2010). One way to counter this problem is to use stratified random sampling to obtain a more demographically representative sample by giving additional weighting to certain sections of the population. Because Maori are less willing to participate in surveys than non-Maori (Towers 2006), Maori were over-sampled in all age groups to increase their participation rate. Similarly, older age groups were over-sampled because older people are less likely to have access to a computer and/or the internet compared with younger age groups.[118]

An application was made to the Electoral Enrolment Centre of New Zealand, under Section 112 of the Electoral Act 1993, for a list of electors stratified by age (18-120 years) and by Maori descent. The New Zealand electoral roll contains the names of all eligible voters (18 years and over) who are citizens or have resided in the country for one year or more. The electoral roll is updated every four years in the year prior to an election. As at 31 March 2007, 96% of all eligible voters were registered on the electoral roll, providing a database that is highly representative of the New Zealand adult population (Towers 2006).

Although sample selection and sample size is often determined by various strategies, in practice, it is usually dependent on the budget available (Amaya-Amaya et al. 2008). In this thesis it would have been preferable to survey a larger sample but the budget would not allow this. However, in order to obtain a sample that was as demographically representative as

[117] This is discussed further in Section 5.4.4.
[118] In 2006 approximately 70% of people aged 64 years or under, lived in households with access to the internet compared with 50% of people aged 65-74 years and 26% of people aged 75 years and over (Ministry of Social Development 2010).

possible, the population was stratified by age, gender, region and ethnicity. Of the 1.5 million names received, 3,315 people were invited to do the survey.

Because non-response to surveys in effect reduces the sample size which can create bias, it is important to achieve as high a response rate as possible. Dillman (2007) suggests that a high response rate can be achieved by using a 'tailored design' approach – surveys should be developed based on a standard set of principles taking into account the survey situation and the sampling population. This helps to create respondent trust thereby encouraging a higher response rate.

Edwards et al. (2009) explored various methods of increasing the response rate to postal or electronic questionnaires. The authors found that some methods are more effective than others. Strategies such as personalising the questionnaire, using simple headings, keeping the questionnaire short, having an interesting topic, providing a monetary or non-monetary incentive and assuring confidentiality all increased the likelihood of questionnaire completion. All of these strategies were employed in this thesis, including the options of donating $1 to a charity or going into a draw to win $200.

It would have been ideal to obtain the email addresses of eligible voters on the electoral roll as the survey was electronic. Unfortunately, as only postal addresses were available, a two-stage approach was necessary. First, I sent a personalised letter to all potential respondents inviting them to contact me to obtain a link to the online survey. Second, when a respondent made contact, I sent him or her an email with a link to the survey.[119] After a minimum of 18 days I sent a reminder letter to those people who had not responded to the initial invitation. I also sent reminder emails to those people who had received the survey link but had not completed the survey. A total of 322 people (a 10% response rate) completed the survey.

5.4.3 Health services researchers' sample

In addition to surveying the general public by snowball and random sampling, a small sample of people who work in the health sector – mainly health services researchers – was invited to complete the survey.

[119] A copy of the invitational letter sent to the random sample is in Appendix 5.6. A copy of the email sent to potential respondents is in Appendix 5.7.

Health services researchers were surveyed primarily because of their familiarity with QALYs. As the prioritisation framework being developed in this thesis could ultimately be implemented within the health sector the use of 'QALYs' as a measure of HRQoL and length of life is appropriate. However, as previously mentioned, most members of the public are unfamiliar with the term 'QALYS' and therefore the use of general categories to describe HRQoL and length of life (i.e. 'small', 'medium' and 'large') is more appropriate for the random sample survey. In contrast, health services researchers *are* familiar with the use of QALYs.

The levels for the criterion 'individual benefit' were changed from 'small', 'medium' and 'large' to '< 1 QALY', '1-3 QALYS' and '> 3 QALYs' respectively and a small sample of health services researchers surveyed (12 respondents). The preferences of people working in the health sector can then be compared with the preferences of the general public, and feedback relating to the survey can be obtained from people who work in the health field.

5.4.4 Sample representativeness

To evaluate the representativeness of the snowball and the random samples, the demographic characteristics of the samples were compared with the demographic characteristics of the New Zealand population using data from Statistics New Zealand.

In Table 5.3 the age, ethnicity, gender, qualifications, region of residency and health insurance status for the snowball and random sample respondents are listed, together with the corresponding statistics for the New Zealand population. In terms of age, compared with the national statistics, the snowball sample has fewer respondents in the over-65 age group and the random sample has fewer respondents in the younger age groups. In both the snowball and random samples, Asian and Pacific Peoples are under-represented, Maori are slightly under-represented and Europeans are over-represented.[120]

In both samples more females than males answered the survey. The random sample closely aligns with the national statistics in terms of qualifications except for the group with no qualifications which is under-represented. The snowball sample has a very high percentage of people with university degrees or equivalent. Similarly, in terms of the region in which

[120] Because some people identify with more than one ethnicity, the percentages for ethnicity in each group sum to more than 100%.

respondents reside, respondents from the random sample live throughout New Zealand whereas respondents from the snowball sample live primarily in the South Island.

Both the random sample and the snowball sample consist of a greater proportion of people with health insurance than the national average.

Table 5.3: Comparison of the demographic characteristics of the snowball and random samples with the New Zealand population

Demographic characteristics		Snowball sample	Random sample	New Zealand Statistics+
Age	18-24	13.5%	7.1%	13.6%
	25-34	21.1%	9.7%	17.0%
	35-44	23.6%	13.0%	18.8%
	45-54	23.6%	23.0%	18.7%
	55-64	14.2%	27.3%	14.6%
	65+	4.0%	19.9%	17.3%
Ethnicity*	Asian	2.2%	3.1%	10.7%
	European	91.3%	90.4%	75.9%
	Maori	8.8%	8.4%	12.6%
	Pacific Peoples	0%	2.2%	6.3%
	Other	2.2%	0.9%	
Gender*	Female	66.5%	59.6%	51.6%
	Male	33.5%	40.4%	48.4%
Qualifications	No qualifications	0.4%	5.9%	18.7%
	Secondary school	12.4%	33.2%	30.3%
	Other post-secondary school qualifications	12.7%	25.8%	29.1%
	University degree or equivalent	74.5%	35.1%	21.9%
Region	North Island	10.3%	71.4%	76.2%
	South Island	89.5%	28.6%	23.8%
	Other	0.4%		
Health Insurance	Yes	40.7%	44.4%	32.0%#

*over 18 years
+Source: Statistics New Zealand (2011)
#Estimate: Health Funds Association of New Zealand (2010)

Table 5.4 compares the demographic characteristics of the respondents in the snowball and random samples with respect to employment, income, household composition, health care usage, serious illness and type of worker. These characteristics cannot be directly compared with national statistics because of the different ways in which the data were collected in the 2006 Census and how the data were collected in this survey.

There is greater variation in terms of employment status and income in the random sample compared with the snowball sample. Respondents in the snowball sample are predominantly employed or students whereas respondents in the random sample fall into a range of employment categories. In terms of income, respondents in the snowball sample earn substantially more than the respondents in the random sample with 31.3% of respondents in the snowball sample earning over $100,000. However, as respondents were given the option to not state their income, it is not possible to compare incomes across the samples or with national statistics. [121]

Household composition in the snowball and random samples is similar. However, there are 11.2% fewer 'couples with no children' in the snowball sample than in the random sample and 7.7% more 'flatmate(s)' in the snowball sample compared with the random sample. In terms of health care usage, 19.3% of respondents in the snowball sample seek health care frequently compared to 30.7% of respondents in the random sample.

Over 65% of respondents in both samples have experienced a serious illness either personally or within their family. More people in the snowball survey have occupations relating to health than those in the random survey.

[121] Respondents were given the option of not stating their income in an attempt to encourage a greater participation rate. This option was included as a result of feedback received from the pilot study. Some respondents said that they felt uncomfortable giving their income and suggested an option be included to decline to answer this question.

Table 5.4: Demographic characteristics of the snowball and random samples

Demographic Characteristics		Snowball sample	Random sample
Employment	Employed full time	60.0%	40.1%
	Employed part-time	16.4%	18.0%
	Looking for work	0%	4.0%
	Homemaker	0.7%	5.9%
	Retired	2.9%	19.6%
	Self-employed	0.4%	4.3%
	Student	19.3%	5.3%
	Other	0.4%	2.8%
Income Group	$20,000 or less	10.9%	9.6%
	$20,001-$30,000	3.6%	10.2%
	$30,001-$50,000	13.1%	12.4%
	$50,001-$70,000	11.3%	18.9%
	$70,001-$100,000	20.0%	16.8%
	$100,001 or more	31.3%	16.8%
	Not given	9.8%	15.2%
Household Composition	Couple/no children	33.8%	45.0%
	Parent(s) with child(ren)	39.3%	33.5%
	Extended family	2.2%	6.2%
	Alone	12.4%	12.1%
	Flatmate(s)	10.5%	2.8%
	Other	1.8%	0.3%
Health Care Usage	Never	0%	0.6%
	Seldom	22.5%	18.0%
	Occasionally	58.2%	50.6%
	Frequently	19.3%	30.7%
Serious Illness	Yes	66.2%	65.5%
	No	33.8%	34.5%
Type of Worker	Medical worker	13.5%	6.8%
	Health-related worker	22.9%	3.4%
	Neither	63.6%	89.8%

With respect to age, qualifications, region of residency and income, the random sample is comparatively more representative of New Zealand's population than the snowball sample. A possible reason for this is that respondents in the snowball sample were encouraged to forward the email to other people whom they thought might be interested in the survey and it is probable that respondents forwarded the email to people similar to themselves in terms of age, income, qualifications and place of residence.

However, while the random sampling method resulted in a more demographically representative sample, the snowball sampling method required less administrative effort and was less costly than the random sampling method, while at the same time producing a fairly large sample (275 respondents) in a short space of time.

Because the random sample is more demographically representative than the snowball sample, and as I would like to explore the relationship between the demographic characteristics of the respondents and their preferences for the criteria, only the data from the random sample survey will be analysed, though the criteria weights for both surveys will be presented in Chapter 6.

5.5 Response rate

As can be seen in Table 5.5, 3,315 letters were sent to potential respondents. A total of 97 respondents were deleted from the survey population due to incorrect addresses, unavailability, death, no English language skills or impairment, leaving a total of 3,218 potential respondents.

The response rate of the random sample is 10% which is the total number of completed surveys divided by the total number of people in the sample. The response rate of 10% in this survey is low compared to other web-based surveys reported in the literature.[122] However, it is difficult to compare response rates between surveys when the topic, format, and structure of the survey are different. In addition, the differences with respect to sample selection, incentives offered and survey administration might account for this discrepancy.

[122] Compounding the low response rate is the number of people who started the survey but did not complete it.

Table 5.5: Survey response rate for the random sample

Survey for random sample		
Letters sent initially	3,315	
Reminder letters sent	3,084	
Letters returned (wrong address)	62	
Deceased	3	
Overseas	4	
No English language skills	14	
Impaired	14	
Total number in sample	97	**3,218**
No computer	45	
Did not wish to participate	4	
Total requests for survey	362	
Survey sent but not started	12	
Survey started but not completed	13	
Couldn't complete because of computer difficulties	7	
Did not wish to complete	1	
Demographic part of survey not completed	7	
Total number of completed surveys		**322**
Response rate (total completed surveys/total number in sample)		**10%**

Response rates to surveys have been declining in most of the industrialised world for several decades (de Leeuw et al. 2002). Nulty (2008) reviewed nine journal articles that examined response rates to paper-based and online surveys for course and teaching evaluation. The online response rates ranged from 20% to 47% with an overall response rate of 33%.

Healey (2005) used a web-based survey to test three of the principles Dillman (2007) suggested to improve questionnaire response rates. An email was sent directly to respondents with a link to a web-based survey followed by a reminder six days later. The response rate was 45%.

A New Zealand study by Smee & Brennan (2000) comparing email, web and mail response rates found a response rate of 12.7% for an email questionnaire and 40.9% for a web-based questionnaire.[123]

Bech & Kristensen (2009) surveyed a random sample of 10,000 50-75 year old individuals and compared the response rate to a postal questionnaire with a letter containing a link to an online version of the same questionnaire. The response rate to the web-based questionnaire was 16.9%.

The means by which the survey was administered may be the key reason for the relatively low response rate. As mentioned earlier, respondents were first sent a letter inviting them to do the survey and asking them to contact me if they were interested. When respondents made contact, I sent them a link to the online survey. This two-step approach could not be avoided as the electoral roll does include email addresses.[124] However, it added a level of complexity to the survey process. Some respondents may have been willing to do the survey if they had been sent a personalised email with a link to the survey but were put off by having to contact me first – thereby reducing the number of respondents.

In addition, the electoral roll was updated prior to the 2006 Census and as the survey was sent in November 2010, some of the names and addresses on the electoral roll would have been incorrect. I became aware of 69 letters with incorrect names and/or addresses but it is likely that there were more than this.

The main concern arising from a relatively low response rate is that the sample is not representative of the larger population. This is known as sampling error. When the sample deviates from the population in terms of characteristics such as age, gender, education and occupation for instance, it is not demographically representative. If the sample is not representative of the population then the survey results from the sample cannot be generalised to the larger population.

[123] With email questionnaires, emails are sent as text messages to potential respondents. The emails can be read, saved, responded to or deleted. With web-based questionnaires surveys are 'posted' on the internet and potential respondents are directed to the website to complete the survey (Brawner et al 2001).

[124] Email addresses of potential respondents can be obtained through 'survey panels'. However, as membership is often by way of rewards programmes, respondents are unlikely to be as demographically representative of the New Zealand population as randomly selected respondents from the electoral roll.

According to Cook et al. (2000), representativeness of the sample is more important than the response rate unless a higher response rate is needed to ensure representativeness. In Section 5.4.4 the demographic characteristics of the random sample were compared with national statistics. There were some differences (the youngest age group, ethnicity, gender, people who were unqualified) but overall the random sample was broadly representative of the New Zealand general public with respect to demographic characteristics.

Another problem associated with a relatively low response rate is that non-respondents may be significantly different from respondents in ways other than demographic characteristics which can lead to biased survey results (Dillman 2007). This is discussed in the next section.

5.6 Non-response bias

Compared to other traditional methods of surveying such as postal, telephone or face-to-face surveys, internet-based surveys have a number of advantages. They are time-saving, cost-saving and provide data in a ready-to-use format. A major disadvantage of internet-based surveys, however, is that they depend on internet access (Blasius 2010) and a respondent's ability or confidence in using a computer.

In 2009, 75% of New Zealand households had access to the internet at home (Statistics New Zealand 2009). It is difficult to know how many people in the sample did not have access to a computer and/or the internet. (I know of 54 respondents from the random sample who did not have access to a computer.[125]) In addition, it is possible that some of the emails sent to respondents were treated as 'junk-mail' or 'spam' and subsequently ignored.

Not having access to a computer or the confidence to complete an online survey contributes to a low response rate, but also creates a non-response bias if the people who are responding to a survey are different (in some relevant way) from those who do not respond. For example, respondents may be more educated or younger than non-respondents.[126]

[125] Most of the 54 respondents either wrote to me or phoned me to let me know that they did not have a computer.

[126] Non-response bias can also refer to 'item non-response' (Bech & Kristensen 2009). This occurs when respondents only partially complete a survey by skipping or ignoring certain questions. In this present study, respondents could not submit the survey until it was completed. Twenty five respondents started the survey but did not complete it. These surveys were eliminated from the analysis and therefore item non-response is not an issue.

It is important to assess whether the people who answer the survey differ in particular ways (e.g. age, interest in health) from those who did not answer the survey to ensure that the estimates for the criteria weights are not biased.

In an effort to ascertain why people did not complete the survey, 35 people who received letters of invitation, were telephoned and asked why they did not complete the survey. The results are reported in Table 5.6. As can be seen in the table, apart from not having a computer, the main reason given for not responding to the survey invitation was that people were either too busy or did not like doing surveys of any kind. Although this sample was small, no-one mentioned that it was the topic of the survey (health prioritisation) that discouraged them from completing it.

Table 5.6: Survey of non-respondents

Telephone survey to check non-response bias	
Total telephone calls to individual non-respondents, randomly selected	35
Male	17
Female	18
Responses:	
No longer at that address	4
Away at the time	1
Too busy	7
Don't do any surveys	8
Don't want to do the survey	3
No computer	9
Don't have the confidence to do the survey	3

According to Sax et al. (2003), a low response rate does not necessarily mean that the sample is unrepresentative of the population. However, estimating non-response bias is challenging as the identity of the non-respondents is often unknown. Sivo et al. (2006) suggests three methods used by Information System researchers to examine non-response error. The first method is to compare the demographic and socioeconomic characteristics of the sample with the wider population (as discussed in Section 5.4.4). The authors argue that if there is no significant difference between the demographics of the respondents and non-respondents,

(assuming that the sample of non-respondents is randomly selected) then there is no response error. The second method is to compare the survey results of the early respondents with those of the late respondents. The results of using this method are discussed later in this section.

The third, and more difficult method, is to contact a group of non-respondents and have them complete the survey using a different mode (e.g. a telephone interview instead of a web-based survey). The differences between respondents and non-respondents can then be assessed. However, the problem with this approach is that non-respondents may be difficult to contact and if a different mode of survey is used responses from both surveys cannot be compared with confidence. As potential respondents in this thesis had already received two letters of invitation to do the survey, it seemed very unlikely that a third letter would elicit a response also bearing in mind the additional expense of sending a third letter.

According to Sivo et al. (2006), late respondents are similar to non-respondents. Therefore if there is a difference between early respondents and late respondents then there will be a difference between respondents and non-respondents. Kypri et al. (2011) investigated non-response bias in a web-based survey by comparing the responses of early respondents with late respondents. Students from 12 New Zealand tertiary institutions received personalised email invitations to complete a web-based survey on health behaviour. Three e-mail reminders were sent. Of the total respondents, 37% replied early, 9% responded late and 54% did not respond. The responses of early respondents were compared with the responses of late respondents with late respondents serving as a proxy for non-respondents. The authors found that late respondents reported more risk behaviour than early respondents. Assuming that the risk behaviour of non-respondents was similar to late respondents, the authors concluded that prevalence of risk behaviour had been substantially underestimated given that the majority of the survey respondents were early respondents.

A similar approach to Kypri et al. (2011) was used in this thesis. The criteria weights and the demographic characteristics of early respondents were compared with those of late respondents to establish whether there were any noticeable differences between the two groups.

As explained earlier, over a period of five days, 3315 letters of invitation were sent to a random sample of potential respondents with 3084 reminder letters sent a minimum of 18 days later. The total number of respondents (322) was divided into two groups with ‘late

respondents' consisting of those respondents who completed the survey after the second reminder was sent.

Independent t-tests[127] were conducted to compare the criteria weights between the early respondents (n = 143) and the late respondents (n = 179). The mean criteria weights for both groups are presented in Table 5.7. No significant differences in the mean criteria weights between the two groups were found. The Levene's test indicates the amount of variability in the groups' mean scores. A Levene's test statistic less than 0.05 indicates that the groups exhibit unequal variance. For all six criteria the significance level of the Levene's Test was greater than 0.05, revealing that the data did not violate the assumption of equal variance.

Table 5.7: Mean criteria weights for early and late respondents

Criteria		Group 1 (n = 143)	Group 2 (n = 179)	Significance (2-tailed)
Only available treatment	Mean	10.31%	10.76%	t (320) = −0.656, p = 0.512
	SD	5.75%	6.46%	
Age	Mean	14.08%	14.22%	t (320) = −0.165, p = 0.869
	SD	7.90%	6.68%	
Societal benefit	Mean	12.17%	12.02%	t (320) = 0.193, p = 0.847
	SD	6.73%	6.29%	
Need	Mean	28.16%	28.56%	t (320) = −0.379, p = 0.705
	SD	8.89%	9.89%	
Individual benefit	Mean	21.69%	22.30%	t (320) = −0.681, p = 0.497
	SD	7.55%	8.38%	
Lifestyle	Mean	13.60%	12.14%	t (320) = 1.664, p = 0.097
	SD	7.55%	8.04%	

Group 1: early respondents; Group 2: late respondents

To test whether the early respondents differed from the late respondents in terms of demographic characteristics, individual chi-square tests for independence[128] were conducted on each of the demographic variables – age, gender, ethnicity, region, income, qualifications, employment status, household composition, experience of a serious illness, health care usage,

[127] The independent t-test compares the mean scores of two or more groups on one continuous variable, in this case the criteria weights.
[128] The chi-square (χ^2) test of independence examines the association between two categorical variables. This test is explained in detail in Chapter 8.

health insurance and type of worker. The results are presented in Table 5.8.[129] A chi-square statistic equal to or less than 0.05 indicates that there is a significant difference between the groups in terms of the demographic characteristics.

Table 5.8: Comparison of demographic characteristics of early and late respondents

Demographic Characteristic	Significance*
Age	χ^2 (5, n = 322) = 11.462, p = 0.043
Gender	χ^2 (1, n = 322) = 0.546, p = 0.460
Ethnicity (3 groups)	χ^2 (2, n = 322) = 0.161, p = 0.923
Region (North Island/South Island)	χ^2 (1, n = 322) = 0.008, p = 0.929
Income	χ^2 (6, n = 322) = 0.676, p = 0.995
Qualifications	χ^2 (3, n = 322) = 1.893, p = 0.595
Employment status	χ^2 (7, n = 322) = 0.692, p = 0.437
Household composition (children/no children)	χ^2 (1, n = 273) = 0.649, p = 0.420
Experience of a serious illness	χ^2 (1, n = 322) = 3.832, p = 0.066
Health care usage	χ^2 (2, n = 320) = 3.061, p = 0.216
Health insurance	χ^2 (1, n = 322) = 0.050, p = 0.822
Type of worker	χ^2 (1, n = 322) = 0.016, p = 0.992

*The Chi-square test significance value uses the Yates Continuity Correction value for 2×2 tables and the Pearson Chi-Square value for tables larger than 2×2.
#The Phi coefficient is used for 2×2 tables and Cramer's V is used for tables larger than 2×2. Cohen's (1988) criteria were used to interpret effect size: 0.10 small effect, 0.30 medium effect and 0.50 large effect.

The age of early respondents is significantly different from the age of late respondents at the 5% level. However, the effect size, indicating the degree of association between the two groups, is small (0.043).[130] There is also a statistically significant difference (at the 10% level) between respondents with respect to 'experience of a serious illness'. The effect size between the two groups is also small (0.05).

In terms of the other demographic characteristics there are no statistically significant differences between early respondents and late respondents. As the demographic characteristics include several health-related attributes such as 'health care usage' and 'health

[129]Several of the demographic variables needed to be collapsed into a smaller number of groups to meet a requirement of chi-square testing, namely that each cell should contain a minimum of five respondents. The 16 regional groups were combined into two groups (North Island/South Island), the seven ethnicity groups were combined into three groups, the six household composition groups were combined into two groups (children/no children) and the four health usage groups were combined into three health usage groups.
[130] Cramer's V (for tables larger than 2×2) is used to determine the effect size. The effect size ranges from 0 to 1 with higher values indicating a stronger association between the two variables. Cohen's (1988) criteria are used to interpret both coefficients, with 0.10 for a small effect, 0.30 for a medium effect and 0.50 for a large effect.

insurance', this implies that in terms of these health characteristics, early respondents are not over-represented in the sample.

Given there is little difference between the early and late respondents (based on the mean criteria weights and the demographic characteristics), and assuming that non-respondents are similar to late respondents (Sivo et al. 2006, Kypri et al. 2011), there does not appear to be a noticeable difference between respondents and non-respondents in this survey.

However, there may be an element of non-response bias due to non-coverage (Couper 2007).[131] Not everyone who was invited to do the survey had access to, or confidence to use, a computer and/or the internet and it is possible that these people differed from the survey respondents in terms of particular demographic characteristics. One way to reduce this potential non-response bias is to provide computer access to those without it, by visiting respondents with a laptop or organising computer access at the University for example. Unfortunately this was outside the scope and budget of this project but it is something to be considered in future research.

As explained previously, although the difference between respondents and non-respondents is considered to be small, the random sample is not completely representative and there may be non-response bias due to a lack of computer/internet access. Therefore analysis of the criteria weights from the random sample should be interpreted carefully.

5.6.1 Generalising the results

According to Dillman (2007), four sources of error need to be considered when generalising the results from a sample to a population: sampling error, non-response error, non-coverage error and measurement error. In this survey stratified random sampling was used to minimise the sampling error. Although the response rate was low at 10%, as discussed in the previous

[131] Couper et al (2007) found in an internet study on health and retirement for over 50 year-olds in the US that non-coverage due to a lack of internet access was of much greater concern than non-response (unwillingness to participate given access). As mentioned in Section 5.4.2, older age groups were over-sampled in this thesis because statistically older people are *less* likely to have access to a computer and/or the internet compared with younger age groups. However, a greater proportion of older people participated in the survey than younger people. Possible reasons for this include: a greater number of older people had access to computers and/or the internet than expected (although the majority of people who contacted me to say that they did not have access to a computer and/or the internet were in the older age groups); receiving a personalised invitation may have encouraged older people to participate compared with younger people; older people may have been more interested in the survey topic than younger people; or older people may have had more time to complete the survey compared to younger people (i.e. lower opportunity costs).

section, the difference between respondents and non-respondents is considered to be small (assuming that late respondents and non-respondents are similar), and therefore the non-response error is assumed to be small. However, as an unknown number of the sample may not have had access to a computer and/or the internet, or the confidence to complete and online survey, there is likely to be an immeasurable non-coverage error.

Measurement error refers to how well a survey captures a respondent's true preferences. Many respondents took the opportunity to comment as they made their decisions (usually to justify why they chose one alternative other the other), thereby providing support for their decisions. In addition, having an easy-to-navigate survey (which is discussed in the next section) contributes to a small, if not non-existent, measurement error.[132] Also, a group of respondents completed the survey twice and the results were compared. The results suggest that the survey accurately captures respondents' preferences. The results of the 'test retest' are discussed in Section 5.8.

However, given the low response rate and the non-coverage error, generalising any results from the sample to the population should be treated with caution.

5.7 Survey design

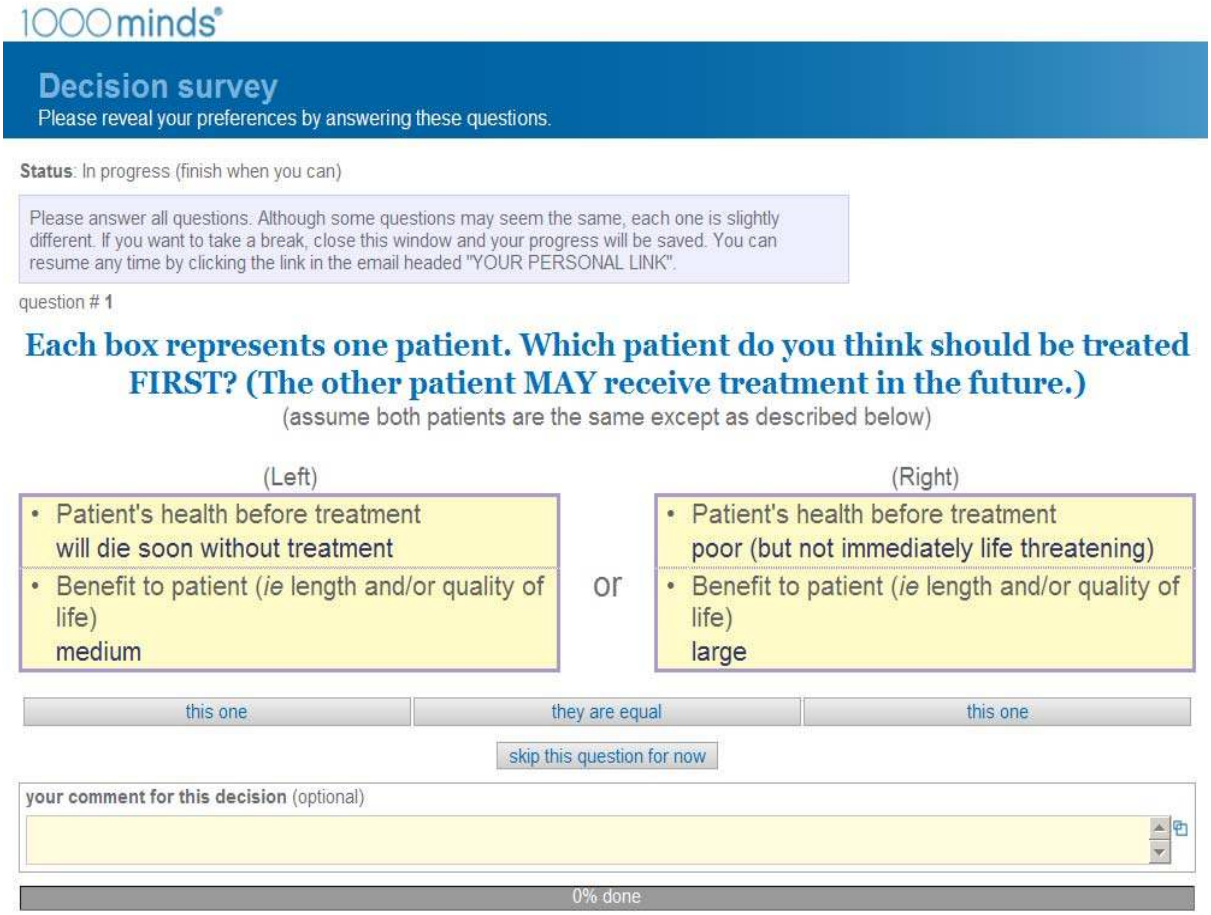
The online survey for this thesis was constructed using 1000Minds software (Ombler & Hansen 2012), Google documents and the PAPRIKA scoring method (Hansen & Ombler 2008) (discussed in Chapter 2). All surveys, regardless of mode, have issues relating to survey design and implementation. Several issues relating to respondent behaviour and survey design are discussed in this section.

5.7.1 Critique of survey

As discussed in Section 5.4.2, potential respondents (from the random sample) were sent a letter inviting them to complete the online survey. When interested respondents made contact, they were sent an email with a *personalised* link to the survey. When a respondent clicked on this link they were taken to the first page of the online decision-survey, as illustrated in Figure 5.2.

[132] A reliable and valid survey keeps measurement error to a minimum. Reliability and validity of the survey are discussed in Section 5.8.

Figure 5.2: Example of a question from the decision survey



Respondents were asked to choose which of two imaginary patients to treat first (left or right). They could also select 'they are equal', 'skip this question for now' and/or write a comment relating to their decision. Respondents could take a break any time they wished by closing the survey and resuming it later by clicking on the personal link in their email. As each respondent completed a survey, the survey data were automatically updated by the software and provided in an easy-to-use format for further analysis.

Using 1000Minds software to create the online survey was very cost-effective due to the minimal set-up and administration costs.[133] Compared to other survey modes such as telephone or face-to-face interviews, numerous respondents in various locations could be contacted simultaneously (by sending an email with a link to the survey) at no extra cost with completion time typically much quicker than that for other surveying methods.

[133] However, obtaining the email addresses of potential respondents incurred some cost as the email addresses of people on the electoral roll were not available; consequently invitational letters and reminder letters had to be sent to potential respondents.

However, several issues were identified in relation to how respondents made their decisions. As can be seen in Figure 5.2 respondents had the option to select 'they are equal' or to skip the question. It is possible that some respondents selected 'they are equal' when really they were 'undecided' rather than 'indifferent'.[134] Respondents may also have clicked any of the options without much thought, to complete the survey quickly. This issue is common to all types of surveys, particularly where there are a large number of questions, or the questions are complex, leading to respondent fatigue (McFadden 2005). However, as respondents considered only two criteria at one time (which minimised the complexity of the task) and the average number of questions respondents answered was 25, this issue is not of major concern.

Another survey response behaviour more common to other modes of surveying, particularly face-to-face interviews, is that some respondents may have answered in a way that they thought they *should* (perhaps from a societal perspective), rather than how they actually felt (McFadden 2005).

It is also possible that some respondents displayed dominant preferences. Respondents with dominant preferences always choose the alternative that contains more of their preferred criterion regardless of the levels of the other criterion. If the preferred criterion is at the same level in both alternatives, then the choice will be made based on the other criterion/criteria (Scott 2002). Even if this occurred, the purpose of the survey is to elicit individual preferences and as such all completed surveys were used in the analysis. As Lanscar (2006, p 809) commented, it is "somewhat paradoxical" that researchers design surveys to elicit consumer preferences "but if the results do not conform to researchers' a *priori* expectations of how preferences 'should' behave", then those 'irrational' responses are deleted which may result in eliminating valid preferences and inducing sample selection bias.[135] Similarly, Bryan & Dolan (2004) argue that it is difficult to justify deleting respondents from a sample because they have dominant preferences, given that the purpose of eliciting their preferences is to inform public policy.

As explained in Chapter 2, when respondents answer a question the software automatically eliminates all other potential questions that are implicitly answered as corollaries of that

[134] One respondent suggested that 'they're equal' could be renamed 'indifferent' as the two patients are not equal if they have different health needs but a person may be 'indifferent' as to who receives treatment first. In the pre-test and pilot study, it was clear that respondents considered 'they are equal' to mean 'indifferent'.

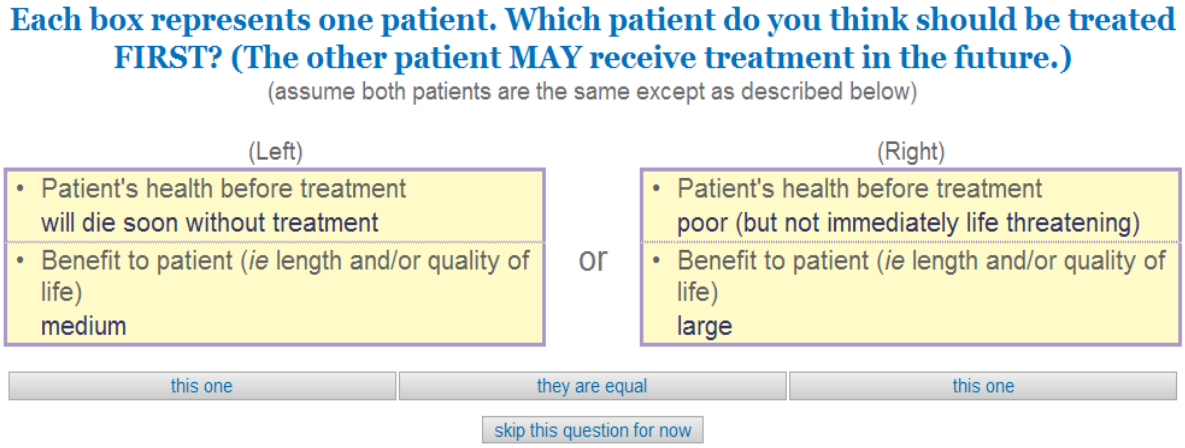
[135] For this reason, outliers (extremely high or low criteria weights) were not eliminated from the data as they represent respondents' preferences.

question. This means that if a respondent inadvertently selected one alternative when their intention was to choose the other, the flow-on effect could result in weights that may not accurately reflect that respondent's preferences. Although this is possible, answering one pairwise question 'incorrectly' is unlikely to greatly affect that person's criteria weights. (As mentioned previously, a 'test re-test' was conducted which supports this assertion. The results of the 'test re-test' are discussed in Section 5.8.) Nevertheless, if a respondent 'incorrectly' answers a question it is highly probable that the affected criteria will be presented again in another question, albeit with different levels. There is also an option to click 'undo last decision' to go back to the previous question (s). Therefore if a respondent realises that they have made a mistake or if they change their mind, they can re-answer the questions. Also as1000Minds software provides weights on the criteria for *every* respondent it is possible to assess whether the criteria weights accurately represent a respondent's preferences by asking them whether the criteria weights correspond with their expectations.

To ensure that respondents correctly understood the meaning of the criteria and/or the levels in the trade-off questions, I sat with the pilot sample respondents when they completed the survey and asked them to verbalise their decision-making. Respondents consistently interpreted the criteria in the same way. The only criterion creating some disparity in interpretation was 'only available treatment'.

As can be seen in Figure 5.3 below, when choosing which patient to treat, respondents were asked to 'assume both patients are the same except as described below'. Some respondents ignored this instruction. For example, one respondent imagined two patients of different ages and gender when the two criteria to be considered were 'need' and 'societal benefit'.

Figure 5.3: Instruction for the trade-off question in the decision survey



Sometimes when respondents think that they need additional information to make a decision, they fill in the 'missing information' themselves (Manksi 1999, Smith 2003). However, the linear additive model assumes that all criteria are independent of each other, that is, that there are no interaction effects.[136] When a respondent chooses between two *pairs* of criteria their decision should not be affected by any other criteria (hence the instruction 'assume both patients are the same except as described'). For example, suppose 'Patient A' is 0-14 years and will receive a small benefit from treatment and 'Patient B' is 55 years or over and will receive a large benefit from treatment. If joint factor independence holds, then a respondent will choose either patient A or patient B, based on the two criteria, independent of all other criteria. However, some respondents in the survey commented that their decision *would* depend on what the other criteria were. For instance, if both patients were classified as 'high need' then Patient A might be chosen, but if both patients were classified as 'low need' then Patient B might be chosen. This implies that joint factor independence does not hold. However, this issue was raised by only two respondents from the pilot sample when ask to discuss the survey. It is impossible to predict whether this occurred with any respondents from the random sample. It seems probable that any effect would be minimal and therefore unlikely to invalidate the criteria weights obtained, particularly when the criteria weights are averaged across the respondents.

[136] This reflects the property of 'joint factor independence' (Krantz 1972, Luce 1992), which was explained in Chapter 2.

5.7.2 Survey format

Respondents were asked about the format of the survey in terms of the design and instructions, and also about how easy or how difficult they found answering the trade-off questions. The results are presented in Figure 5.4.

Figure 5.4: Survey question relating to the survey format and difficulty of questions

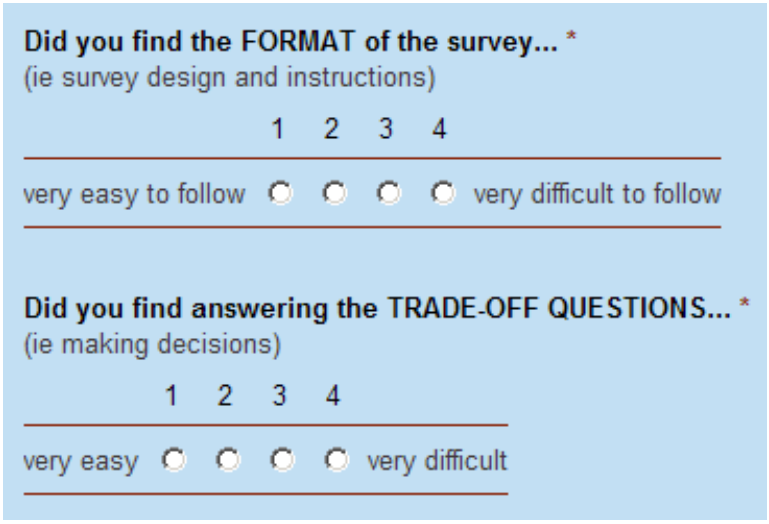


Table 5.9 displays a summary of the results. A large proportion of the respondents from both the snowball and random samples found the survey format, in terms of the survey design and instructions, easy or very easy to follow (81.8% of the snowball sample and 84.5% of the random sample). Only 18.2% of the snowball sample and 15.5% of the random sample found the format of the survey difficult or very difficult to follow.

Table 5.9: Results of question in regard to survey format and difficulty of questions

Survey format and difficulty		Snowball Survey (*N*=275)	Random Survey (*N*=322)
Survey Format	Very easy	50.5%	49.1%
	Easy	31.3%	35.4%
	Difficult	17.1%	12.7%
	Very difficult	1.1%	2.8%
Decision Difficulty	Very easy	4.4%	8.1%
	Easy	29.1%	32.6%
	Difficult	44.7%	44.1%
	Very difficult	21.8%	15.2%

It is encouraging that a high percentage of respondents judged the survey design and instructions easy to follow. The easier a survey is to complete, the more likely it is that respondents will complete it.

A criticism relating to surveys in general is that respondents are often not given much time to consider questions before answering them. According to Dolan (1999, p 916), "if the considered opinions of the general public are required, surveys that do not allow respondents time or opportunity for reflection may be of doubtful value." In this survey, respondents were given the option to close the survey and to finish it at a later date by re-clicking on the survey link. This allowed respondents to think about the questions for as long as they liked and to complete the survey in their own time which may have contributed to the high number of people who found the survey format easy to follow.

In terms of the ease with which respondents answered the trade-off questions, proportionately more respondents found the questions difficult[137] to answer (66.5% of the snowball sample and 59.3% of the random sample) than easy to answer (33.5% of the snowball sample and 40.7%). It is not surprising that up to two-thirds of respondents found it difficult to answer the trade-off questions. Having to choose between two patients who both need treatment requires some thought. However, as discussed in Chapter 2, choosing between pairs with just two criteria is a much easier task than choosing between pairs with more than two criteria as it minimises the level of complexity, reduces respondent fatigue and, ultimately, leads to more robust results.

In the next section the validity and the reliability of the survey are discussed along with the results of a 'test re-test'.

5.8 Validity and reliability of the survey

As discussed in Section 5.6.1, when generalising the results from a sample to a population four sources of error need to be considered including the sampling error, non-response error, non-coverage error and measurement error (Dillman 2007). In addition, the validity and the reliability of the survey also need to be assessed.

[137] In this context, 'difficult' refers to the 'cognitive' difficulty of the survey (i.e. choosing which patient to treat) as opposed to the 'practical' difficulty (i.e. understanding how to complete the online survey).

Validity refers to how well a survey measures what it intends to measure. It can be investigated by examining the 'face', 'content' and 'construct' validity of the survey (Radhakrishna 2007).

A MCDA survey has face validity if all the appropriate criteria and levels are included, with no relevant criteria being excluded or irrelevant criteria being included. In this survey, focus groups were conducted, a literature search undertaken and health specialist advice sought to ensure that the relevant criteria and levels were included in the survey. To have content validity a survey needs to be well-constructed and to use wording that is easily understood by the respondents. A pre-test was conducted and a pilot sample recruited to test the face and content validity of the survey. Some minor changes were made to the survey. As discussed in the previous section the end result was an easy-to-follow survey that was clearly understood by most respondents, encompassing all relevant criteria (except for the criteria that were explicitly left out to be considered later, e.g. cost).

Construct validity, in the context of this thesis, refers to how well the survey captures the preferences of the respondents. A simple way to check construct validity is to give respondents the results of their survey and ask them if the criteria weights accurately reflect their preferences. According to the respondents in the pre-test sample and the pilot sample, their criteria weights closely aligned with their preferences.

Another way to check construct validity is to use the health vignette rankings from the focus groups as a 'gold standard' against which the health vignette rankings of the random sample group can be compared.[138] This approach is suitable when the focus groups consist of participants who have some knowledge and experience of prioritising health treatments. However, the membership of the focus groups was mixed. Also, different terminology was used in the focus groups compared with the decision survey. For example, 'number of patients' was included in the vignette descriptions used in the focus groups in contrast to the decision survey where respondents were asked to choose between two hypothetical patients. Therefore the rankings from the focus groups are not directly comparable with the rankings from the sample groups. However, for illustrative purposes the rankings from the focus groups are compared with the rankings from the three samples (random, snowball and health services researchers) in Chapter 6.

[138] The vignette rankings for the random sample are obtained by categorising each vignette according to the criteria and levels and applying the weights obtained from the survey.

Reliability refers to the extent to which a survey can be interpreted consistently across different situations. The easiest and most common way to assess reliability is to test the same group of people twice, with a suitable time gap. If the results from both surveys are similar then the survey can be considered reliable. This is known as 'test-retest' reliability (Field 2009).

When the time period between the two tests is short (for example, a few days) there is a risk of the 'practice effect'. This is when respondents try to alter the way they answer the questions because they are familiar with the survey. However, when the time period between the two tests is more substantial (for example, a month) respondents are less likely to remember the questions they have been asked in the first survey, but on the other hand they may be influenced by other factors. For example, respondents may become more aware of news items relating to health care prioritisation which could influence the way they answer the questions in the second survey. Ideally the time period between administering the two surveys needs to be long enough for respondents to have had a break but not so long that their decisions are affected by conversations, media or other influences. Changing the order of the questions also helps to minimise the 'practice effect'.

There is no standard rule for determining an appropriate gap between surveys. Instead, it is important to be aware of the factors that could influence survey results and adjust the time period accordingly (Hayes 2008). As this is the first time a 1000Minds survey has been completed twice by the same group of people, there is no precedent for an appropriate time period for this type of survey. A brief search of the literature suggests that approximately two weeks is a common time period between tests, regardless of mode. Badland & Schofield (2006) tested the reliability of a telephone survey measuring transport-related physical activity in New Zealand adults; the test re-test period was 7-10 days. Schlademann et al. (2008) tested the reliability of a paper-based questionnaire relating to back pain severity in a German population sample, using a test re-test period of two weeks. Ferreira & Veiga (2008) tested the reliability of a paper-based questionnaire for screening adolescents at risk of eating disorders, using a test re-test period of 15 days. The reliability of a web-based survey relating to alcohol measures was tested by Miller et al. (2002) using a test-retest period of one week. Bae et al. (2010) used a period of two weeks to test the reliability of a web-based questionnaire on youth risk behaviour.

To test reliability of the decision survey, a convenience sample of 29 respondents was recruited to complete the survey twice, approximately 12 days apart. 1000Minds software keeps a record of every trade-off question a respondent directly answers. Table 5.10 presents a summary of the number of questions directly answered by the respondents in each survey. The total number of questions directly answered is almost identical in both surveys. For example, the minimum number of questions answered in both surveys is 14, the maximum number answered is 37 in the first survey and 38 in the second survey, and the average number of questions in both surveys is close to 23 questions. The number of questions answered as 'they're equal' is the same in both surveys.[139]

Table 5.10: Summary of number of questions directly answered by respondents in test re-test

Respondents (*N*=29)	Minimum	Maximum	**Mean**
No. of questions answered in 1st survey	14	37	**23.3**
No. of questions answered in 2nd survey	14	38	**23.4**
No. of "they're equal" in 1st survey	0	7	**2.8**
No. of "they're equal" in 2nd survey	0	7	**2.9**
Time between surveys (days)	6	35	**12.2**
No. of identical questions	2	15	**7.4**
Identical questions answered the same	1	12	**4.9**
Identical questions answered similarly	0	4	**0.9**
Identical questions answered opposite	0	5	**1.6**

Because1000Minds software automatically changes the order of the trade-off questions (which helps to reduce the 'practice effect') the *first* trade-off question presented to one respondent could be completely different to the *first* trade-off question presented to another respondent. The second question a respondent is presented with depends on how they

[139] A detailed list of the number of questions directly answered by every respondent, including identical questions is in Appendix 5.8.

answered the first question and so on.[140] However, as 1000Minds keeps track of all the questions respondents directly answer, the total number of *identical* trade-off questions presented to a respondent in *both* surveys can be determined. On average, 7.4 identical trade-off questions were presented in both surveys. Of these, an average of 4.9 questions were answered in exactly the same way, an average of 1.6 questions were 'strongly' inconsistent (i.e. the first time the question was answered one patient was ranked ahead of the other and the second time the question was answered, the ranking was reversed), and an average of 0.9 were 'weakly' inconsistent (i.e. the first time the question was answered one patient was ranked ahead of the other; the second time the question was answered, both patients were equal or vice versa).

To gauge whether a practice effect existed, after each respondent completed the second survey I asked them whether they consciously tried to answer the questions a certain way because they were familiar with the survey. All respondents said that their decisions were not influenced by having completed the survey previously, although some respondents said it was quicker to answer because they were familiar with the format.

To assess how similarly respondents answered the trade-off questions in both surveys, a paired samples t-test was conducted. The results are explained in the next section.

5.8.1 Test-retest results

A paired samples t-test [141] is used when data are collected from one group of respondents at two different points in time and compared. To assess whether there is a statistically significant difference in the mean criteria weights between the first and second surveys, a paired samples t-test was used. The criteria weights for both surveys are presented in Table 5.11.

[140] As explained in Chapter 2, when a respondent answers a question the software automatically eliminates all other potential questions that are implicitly answered as corollaries of that question (by applying transitivity).
[141] The paired samples t-test is also known as the 't-test for dependent samples' or 't-test for repeated measures'.

Table 5.11: Mean criteria weights for test re-test sample

Criteria (*N*=29)	Survey 1		Survey 2	
	Mean	SE	Mean	SE
Patient's health before treatment (health status)	27.77%	1.61%	29.49%	1.36%
Benefit to patient (*ie* length and/or quality of life)	23.70%	1.21%	22.06%	1.29%
Illness or injury NOT caused by lifestyle choices	14.40%	1.30%	12.68%	1.59%
Age of patient	12.56%	1.14%	12.80%	1.41%
Benefit to others (*eg* family or society)	11.48%	1.06%	12.79%	1.20%
Only treatment option available for this patient	10.10%	0.83%	10.19%	1.09%

Several assumptions need to be met before conducting a paired samples t-test. The assumptions are similar to those used for one-sample t-tests but relate to the *differences* between the weights. The data need to be at an interval or ratio level, the observations need to be independent from each other and the differences between the weights need to be normally distributed.[142] The criteria weights are at a ratio level, respondents answered both surveys unaware of other respondents in the sample, and the differences in the criteria weights across individuals are normally distributed as indicated in Table 5.12 (a p-value greater than 0.05 in the Shapiro-Wilk test indicates that the data are normally distributed[143]).

Table 5.12: Test of normal distribution of the differences between the means

	Shapiro-Wilk		
	Statistic	df	p-value
Difference Age	0.965	29	0.444
Difference Lifestyle	0.968	29	0.519
Difference Individual	0.968	29	0.517
Difference Society	0.979	29	0.807
Difference Need	0.974	29	0.677
Difference Alternative	0.972	29	0.606

[142] Data that are normally distributed are represented graphically by a typical bell-shaped curve with the highest frequency of data in the middle and smaller frequencies of data evenly distributed towards the edges.
[143] The Shapiro-Wilk Test determines whether a distribution of scores is significantly different from a normal distribution. Because this test is affected by large samples in which small deviations from normality yield significant results, it is recommended for sample sizes less than 50.

With a paired samples t-test, differences in the demographic characteristics of the respondents such as gender, age and income are controlled for by having the same group of respondents answer the survey twice. Paired differences for each criterion are obtained by subtracting the first score from the second score for each criterion. The paired samples t-test determines whether the average differences between the two means on each criterion are significantly different from zero. If the p-value is less than 0.05 there is a statistically significant difference between the criteria means across the two samples.

Table 5.13: Paired samples t-test results

		Paired Differences (*N*=29)							
					95% CI of Difference				
		Mean	SD	SE Mean	Lower	Upper	t	df	p-value
Pair 1	Only available treatment & Only available treatment 2	−0.09%	5.92%	1.10%	−2.34%	2.17%	−0.08	28	0.938
Pair 2	Age & Age 2	−0.24%	7.33%	1.36%	−3.03%	2.55%	−0.18	28	0.862
Pair 3	Societal benefit & Societal benefit 2	−1.32%	6.58%	1.22%	−3.82%	1.18%	−1.08	28	0.290
Pair 4	Need & Need 2	−1.72%	9.58%	1.78%	−5.37%	1.92%	−0.97	28	0.341
Pair 5	Individual benefit & Individual benefit 2	1.65%	8.81%	1.64%	−1.70%	4.99%	1.01	28	0.323
Pair 6	Lifestyle & Lifestyle 2	1.72%	7.80%	1.45%	−1.24%	4.69%	1.19	28	0.244

As can be seen in Table 5.13, all significance values are greater than 0.05 indicating that there are *no* statistically significant differences in the mean criteria weights between the two surveys. The mean differences between each set of criteria are shown in the first column, with −0.09 being the smallest mean difference for the criterion 'only available treatment' and −1.72 being the largest mean difference for the criterion 'need' and 1.72 for 'lifestyle'. (The standard deviation, standard error and confidence intervals refer to the mean differences in the criteria weights.)

A Wilcoxon signed ranks test was also conducted, which is the non-parametric equivalent of a paired samples t-test.[144] The Wilcoxon test converts the criteria weights to ranks then

[144] Non-parametric tests do not rely on estimating parameters and are used when the data are ranked or ordered. Non-parametric tests compare medians whereas parametric tests compare means and are therefore less sensitive than parametric tests at detecting effects.

compares the ranks between surveys. A significance level of 5% is commonly used as a cut-off, but in general, the lower the p-value, the more likely that the criteria scores between the surveys differ. As indicated by the p-values in Table 5.14, there are *no* statistically significant differences between the criteria weights in the first and second surveys.

Table 5.14: Wilcoxon Signed Ranks test results

Wilcoxon Signed Ranks test	Z	p-value
Only available treatment & Only available treatment 2	−0.205[a]	0.837
Age & Age 2	−0.132[a]	0.895
Societal benefit & Societal benefit 2	−1.114[a]	0.265
Need & Need 2	−0.876[a]	0.381
Individual benefit & Individual benefit 2	−0.854[b]	0.393
Lifestyle & Lifestyle 2	−0.865[b]	0.387

a. Based on negative ranks; b. Based on positive ranks; c. Wilcoxon Signed Ranks Test

Although the sample used to test reliability was small (*N*=29) and was not representative of the general public, the results are promising. Both the parametric and non-parametric tests reveal that there are no differences significant at the 5% level (or even 20% level) between the criteria weights in the first and second samples. This suggests that the online survey measured the preferences of the respondents consistently.

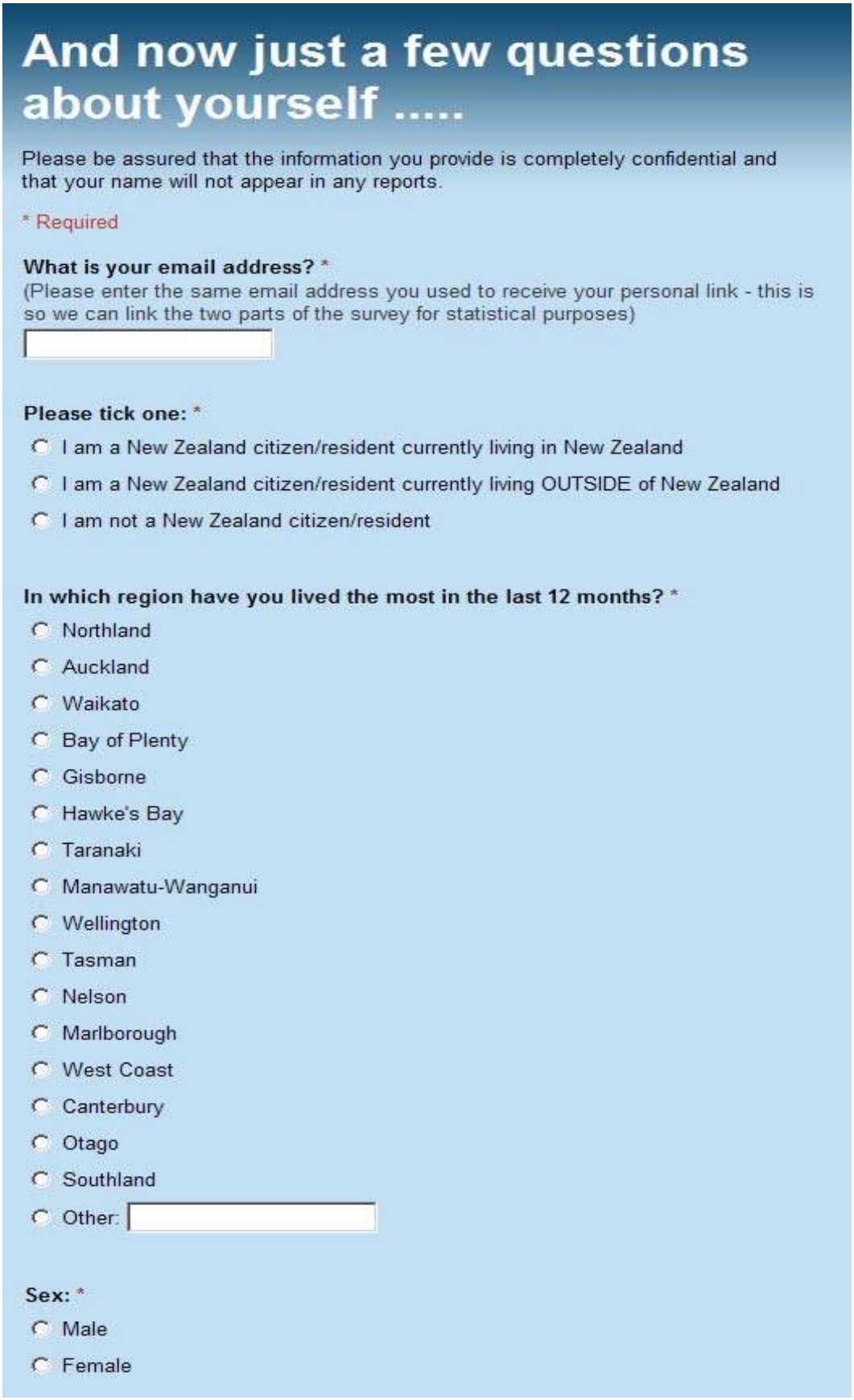
5.9 Conclusion

To elicit the preferences of the general public with respect to the six criteria (‘need’, ‘individual benefit’, ‘societal benefit’, ‘age’, ‘no alternative treatment’ and ‘lifestyle’) a decision survey was developed using 1000Minds software, Google documents and the PAPRIKA scoring method. Three sample groups were surveyed.

Issues relating to demographic representativeness, the response rate and the survey design are discussed in this chapter. As the response rate for the random sample is fairly low and there is the possibility of a non-coverage error, generalising the results from the random sample to the population should be done with caution.

The chapter concluded with an evaluation of the validity and reliability of the survey. The results suggest that the online decision survey was an effective way to elicit the preferences of the general public for criteria that could be used in health care priority-setting.

Appendix 5.1: Demographic part of the online decision survey



And now just a few questions about yourself .....

Please be assured that the information you provide is completely confidential and that your name will not appear in any reports.

* Required

What is your email address? *
(Please enter the same email address you used to receive your personal link - this is so we can link the two parts of the survey for statistical purposes)

Please tick one: *

- I am a New Zealand citizen/resident currently living in New Zealand
- I am a New Zealand citizen/resident currently living OUTSIDE of New Zealand
- I am not a New Zealand citizen/resident

In which region have you lived the most in the last 12 months? *

- Northland
- Auckland
- Waikato
- Bay of Plenty
- Gisborne
- Hawke's Bay
- Taranaki
- Manawatu-Wanganui
- Wellington
- Tasman
- Nelson
- Marlborough
- West Coast
- Canterbury
- Otago
- Southland
- Other:

Sex: *

- Male
- Female

Age: *

- ○ 18-24
- ○ 25-34
- ○ 35-44
- ○ 45-54
- ○ 55-64
- ○ 65 yrs or over

Ethnicity: *
(You can tick more than one box or write in other, eg Dutch, Japanese, Tokelauan)

- ☐ New Zealand European
- ☐ Maori
- ☐ Samoan
- ☐ Cook Island Maori
- ☐ Tongan
- ☐ Niuean
- ☐ Chinese
- ☐ Indian
- ☐ Other:

Highest qualification: *

- ○ No qualifications
- ○ Secondary school
- ○ University degree or equivalent
- ○ Other post secondary school qualification

Work status: *
(tick the one that best describes you or write in other)

- ○ Employed full-time
- ○ Employed part-time
- ○ Looking for work
- ○ Homemaker
- ○ Student
- ○ Retired
- ○ Other:

From all sources of income (excluding loans) what was your total HOUSEHOLD income in the last 12 months? *
(before tax)

- ○ $20,000 or less
- ○ $20,001 - $30,000
- ○ $30,001 - $50,000
- ○ $50,001 - $70,000
- ○ $70,001 - $100,000
- ○ $100,001 or more
- ○ I don't want to answer this question

Who lives with you? *
(Tick all boxes that apply to you)

- ☐ My partner/wife/husband
- ☐ My mother/father
- ☐ My brother(s)/sister(s)
- ☐ My son(s)/daughter(s)
- ☐ Other family members (grandparents, nieces/nephews...)
- ☐ My flatmate(s)
- ☐ I live alone
- ☐ Other:

How often do you and your family members use health care services? *

- ○ Never
- ○ Seldom (eg 1 GP visit/year...)
- ○ Occasionally (eg 3 GP visits/year, occasional medication...)
- ○ Frequently (eg visit GP at least once/month, on prescription medication...)

Have you or a family member experienced a serious illness? *

- ○ Yes
- ○ No

Are you.. *

- ○ A MEDICAL health care worker such as a nurse, doctor, physio...
- ○ A worker in a HEALTH RELATED field such as a health analyst, health researcher...
- ○ Neither

Do you have private health insurance? *

- Yes
- No

Did you find the FORMAT of the survey... *
(ie survey design and instructions)

	1	2	3	4	
very easy to follow					very difficult to follow

Did you find answering the TRADE-OFF QUESTIONS... *
(ie making decisions)

	1	2	3	4	
very easy					very difficult

If you have any comments you would like to make about this survey or on health prioritisation in general, please write them here.

Would you like to receive a summary of the results?
(They should be available mid-2011)

- Yes
- No

That's it! THANK YOU for taking part in this survey. Please tick one of the following:
THEN CLICK THE SUBMIT BUTTON BELOW. (You will then see a blank screen and can close the window.)

- Donate $1 to the Christchurch Earthquake Relief Fund on my behalf
- Donate $1 to the SPCA on my behalf
- Donate $1 to Oxfam New Zealand on my behalf
- Put me into a draw to win $200

Submit

Appendix 5.2: Email sent to snowball sample

Subject: An invitation from Trudy Sullivan

Hi there

Limited dollars - unlimited patients - who would YOU treat first?

This is an invitation to participate in an online decision survey relating to health care prioritisation in New Zealand.

The aim of this survey is to find out the importance New Zealanders place on some of the factors that could be considered when prioritising health services, for example the benefit to a patient of receiving treatment or the part lifestyle has played in a person becoming ill. I would like to find out what YOU think.

The information you provide will be treated with strict confidentiality and your anonymity will be ensured.

The survey consists of a series of hypothetical choices where you will be asked to trade-off one characteristic of health care with another by choosing between two patients.

Below is a link to the on-line survey. When you click this link you will be taken to a webpage where you will be asked for your email address. You will then be sent a **personal link** to the online decision survey.

http://engine.1000minds.com/xStart.aspx?p=fuj&x=cpyj&c=2jmdwq5gw4k3dqed6846x&s=2

If you know of anyone who may be interested in doing this survey, please forward **THIS** email to them and they can get their own personal link by clicking on the link in this email. **Only one survey can be completed per one email address.** The survey closes on 15 November.

If you have any questions or comments email trudy.sullivan@otago.ac.nz or ring me on 479 8134.

What you think matters. Thank you for taking part in this research.

Trudy Sullivan
Department of Economics
University of Otago

Appendix 5.3: Newspaper article published in the *Gore Ensign* on 15/10/10

Treatment priorities focus of survey

By SANDY EGGLESTON

HOW to decide which patients should be given priority treatment in a public health system that has limited funding is a question that should be discussed by New Zealanders, University of Otago student Trudy Sullivan says.

Ms Sullivan is completing a research project for her doctorate to find out what New Zealanders think about which groups of patients should have priority in receiving publicly funded treatment. She would like as many people as possible to take part in an online survey to share their opinion.

Ms Sullivan said because many people complained there was not enough funding for the public health system, there should be clear guidelines as to which patients should have priority.

"People complain 'why am I on a waiting list?' or 'why am I not on it?' and 'how do I know if I can get this treatment?'."

"There needs to be some clear, transparent way to prioritise treatments," she said.

The survey asked people to choose between different criteria that could be used to determine which patient should receive treatment first.

The criteria are age, lifestyle choices a person has made, how much the treatment will help the patient, how much the community would benefit from the person being treated, whether there is an alternative treatment and the patient's health.

Ms Sullivan said the survey would help her discover what criteria New Zealanders thought were important in making health-treatment decisions.

The survey took about 10 minutes to complete and people interested in taking part could email her to receive a computer link to a survey form, Ms Sullivan said.

❑ To receive a computer link to take part in the survey, email trudy.sullivan@otago.ac.nz

Appendix 5.4: Newspaper article published in the *Otago Daily Times* on 23/10/10

4 GENERAL

Treatment-priority survey

By ELSPETH McLEAN

IF you had to rank 14 health treatments ranging from hip replacements to drugs for erectile dysfunction according to their benefit or value to society, how would you do it?

That task was given to six focus groups by researcher Trudy Sullivan of the University of Otago's economics department as part of her PhD study.

She is trying to find out the importance New Zealanders place on some of the factors that could be considered when setting priorities for health spending where demands exceed a limited budget.

The focus groups — nurses, health policy makers, over-65-year-olds, public health researchers, GPs, and Maori health service providers — were given some information about 14 different treatments (see fact box) and asked to rank them without considering the cost.

The information included prevalence of the condition or illness and the side-effects from the treatment.

Ranking list

How would you rank these treatments according to their benefit to society?

- Service for postnatal depression.
- Providing hand sanitiser in primary schools.
- Statins for patients at high risk of cardiovascular disease.
- Methadone for opioid addiction.
- Vaccine to prevent cervical cancer.
- Growth hormone for the rare Prader-Willi syndrome.
- Oral drugs for erectile dysfunction.
- Antiretroviral drugs for HIV.
- Invitro fertilisation treatment.
- Abatacept for last-line treatment of rheumatoid arthritis.
- Dialysis for end-stage kidney disease.
- Imatinib mesylate for chronic myeloid leukaemia.
- Hip replacements.
- Position emission tomography (PET scan).

Ms Sullivan said the groups eventually ranked the treatments by majority consensus with four out of the six putting oral drugs for erectile dysfunction at the bottom and the same number ranking statins for patients at high risk of cardiovascular disease at the top.

Ms Sullivan said while there was reasonable agreement about the top and bottom priorities, people were divided over ranking of such things as hand sanitiser in schools.

Treatment for illnesses considered to have a lifestyle element, and late-stage treatment for leukaemia and kidney disease were lowly ranked.

Ms Sullivan said it was not the ranking that mattered, but getting people to talk about how they made their choices.

From that she gleaned six factors which those in the groups thought should be included in a prioritisation process — age; benefit to patient (length and/or quality of life); benefit to others (family or society); whether there were alternative treatments; how sick patients were before the treatment; and whether the illness or injury had been caused mainly by lifestyle choices.

What weight people give to these factors is now being tested by Ms Sullivan in an online survey in which people are given a series of scenarios where they must choose between two hypothetical patients.

Ms Sullivan hopes her thesis, which she hopes to have completed in 15 months, may be used by policy makers to aid decision making and that it might lead to further research in the area.

Her survey, which takes 10 to 20 minutes to complete, is open for anyone to complete until November 15.

Later, Ms Sullivan will also be randomly selecting 1500 people from electoral rolls and inviting them to complete the survey, something she expects may yield 200 to 300 participants.

She will be comparing the results of the random group with the self-selected groups.

About 200 people from the latter group had completed the survey so far.

● Anyone wishing to complete the survey can contact Ms Sullivan at *trudy.sullivan@otago.ac.nz*

elspeth.mclean@odt.co.nz

Appendix 5.5: Email sent from 1000Minds to snowball sample respondents

Subject: Your personal link

Dear {fullname}

This is **your personal link** to the online decision survey. If you want to take a break while completing the survey, you can close the survey window (progress will be saved) and resume at any time by clicking on the following link.

{url}

This link is uniquely tied to the survey and your email address so please **DO NOT FORWARD** this message to other people.

You can expect to answer around 25 trade-off questions. You might find answering some of them tricky - that's fine! I want to know what **YOU** think, so relax, sit back and click away.

Please complete the survey by 15 November. If you have any questions or comments email {reply-to} or ring me on 479 8134.

Thank you for taking part in this research.

Trudy

Appendix 5.6: Invitational letter sent to random sample

8 November 2010

{address}

Dear {name}

Huge demand for health services. Not enough money.
How do we decide who gets treated and who doesn't?

Do you want to have a say in how health treatments could be prioritised?

I want to know what YOU think!!!!

This is an invitation to participate in an online decision survey relating to health care prioritisation in New Zealand.

My name is Trudy Sullivan. I'm in the Economics Department at the University of Otago. I'm investigating what could be included in a prioritisation process by finding out how important some health care characteristics are to New Zealanders, such as the benefit to a patient of receiving treatment or the part lifestyle has played in a person becoming ill. I want to find out what YOU think.

You have been randomly selected from the electoral roll to take part in this survey. Any information you provide is totally confidential and your name will not appear in any reports. Your participation in this survey is completely voluntary - you can withdraw at any stage.

The survey consists of a series of hypothetical choices where you will be asked to trade-off one characteristic of health care with another by choosing between two patients. There will be around 24 trade-off questions to answer so it will take around 10-15 minutes to complete. You will need an email address and access to a computer with internet.

I'd be very grateful if you would take part in this research. To obtain **your personal link** to the survey, please email me at trudy.sullivan@otago.ac.nz.

As a small thank you for participating, at the end of the survey you can choose to go into a draw to win $200, OR to select one of three charities to which we will donate $1. Please complete the survey by **30 November 2010**.

If you have any questions please ring me on (03) 479 8134 or email trudy.sullivan@otago.ac.nz.

What YOU think matters. Thank you for taking part in this research.

Yours sincerely

Trudy Sullivan

Appendix 5.7: Email sent from 1000Minds to random sample respondents

Subject: Your personal link

Dear {fullname}

PRIORITISING HEALTH CARE - WHAT IS IMPORTANT TO NEW ZEALANDERS?

THANK YOU for agreeing to take part in this research project. This is **your personal link** to the online decision survey (ie click on the blue link below and it will take you straight to the survey). If you want to take a break while completing the survey, you can close the survey window (progress will be saved) and resume at any time by clicking on the link.

{url}

This link is uniquely tied to the survey and to your email address so **DON'T FORWARD** this message to other people.

Some questions you might find tricky - that's fine! There's no right answer. I want to know what **YOU** think, so relax, sit back and click away.

Please complete the survey **as soon as you can**. If you have any questions or comments email {reply-to} or ring me on (03) 479 8134.

Thank you for taking part in this research.

Trudy

Appendix 5.8: Number of questions directly answered by respondents in test re-test

Test Re-test Sample (n = 29)

Respondent	No. of questions answered in 1st survey	No. of "they're equal" in 1st survey	No. of questions answered in 2nd survey	No. of "they're equal" in 2nd survey	Time between surveys (days)	No. of identical questions	Identical questions answered the same	Identical questions answered similarly	Identical questions answered opposite
1	20	5	21	5	24	3	2	1	0
2	29	0	24	0	25	10	8	0	2
3	26	2	19	2	7	8	7	0	1
4	29	1	21	6	17	9	4	3	2
5	26	1	22	0	7	8	4	4	0
6	26	1	24	1	14	7	4	0	3
7	23	3	26	4	8	7	7	0	0
8	24	0	22	0	7	6	4	0	2
9	26	1	38	2	8	14	10	1	3
10	33	0	35	0	7	15	12	0	3
11	29	0	20	0	7	8	6	0	2
12	28	0	18	5	7	7	5	2	0
13	37	0	28	0	35	15	10	0	5
14	25	0	34	0	13	12	9	0	3
15	23	0	26	0	14	11	6	0	5
16	27	1	25	5	15	7	6	0	1
17	16	6	19	5	6	8	6	1	1
18	15	5	21	5	11	3	2	0	1
19	21	4	21	0	10	6	3	2	1
20	21	3	27	0	13	6	1	1	4
21	16	5	21	6	7	4	3	1	0
22	20	6	20	4	16	6	3	2	1
23	14	7	14	6	8	2	0	2	0
24	20	6	22	7	7	6	4	1	1
25	17	3	25	7	8	6	3	2	1
26	19	6	18	4	16	5	3	2	0
27	24	3	26	1	7	8	6	1	1
28	20	6	19	5	14	3	3	0	0
29	22	5	23	5	17	5	2	1	2
Minimum	14	0	14	0	6	2	1	0	0
Maximum	37	7	38	7	35	15	12	4	5
Average	**23.31**	**2.76**	**23.41**	**2.93**	**12.24**	**7.41**	**4.93**	**0.93**	**1.55**

~ Chapter 6 ~

Respondents' preferences for the six criteria

6.1 Introduction

In this chapter the criteria weights obtained from the three samples are discussed and compared with results from comparable studies, and a summary of the many comments received from respondents throughout the survey process is presented.

6.2 Preferences

The mean weights for the six criteria ('need', 'individual benefit', 'age', 'lifestyle', 'societal benefit' and 'only available treatment') for the three sample groups (random sample, snowball sample and health services researchers' sample) are presented in Table 6.1. The weights represent the preferences of the respondents with respect to the relative importance of the six criteria. Also included in the table are the relative rankings of the criteria for each of the samples.

Table 6.1: Criteria weights for the three samples

Criteria	Random sample (*N*=322)	Snowball sample (*N*= 275)	Health services researchers' sample (*N*=12)
Need	**28.4%**	**29.7%**	**31.5%**
Std. Deviation	9.4%	10.2%	9.1%
Rank	1st	1st	1st
Individual benefit	**22.0%**	**24.6%**	**27.9%***
Std. Deviation	8.0%	8.7%	7.0%
Rank	2nd	2nd	2nd
Age	**14.2%**	**13.1%**	**11.3%**
Std. Deviation	7.2%	7.1%	9.7%
Rank	3rd	3rd	4th
Lifestyle	**12.8%**	**11.0%**	**5.8%**
Std. Deviation	7.9%	7.5%	6.1%
Rank	4th	5th	6th
Societal benefit	**12.1%**	**12.0%**	**13.1%**
Std. Deviation	6.5%	6.9%	8.1%
Rank	5th	4th	3rd
No alternative treatment	**10.5%**	**9.6%**	**10.4%**
Std. Deviation	6.2%	6.3%	5.9%
Rank	6th	6th	5th

*In the health services researchers' survey, the levels for benefit to patient were "small (<1 QALY)", "medium (1-3 QALYs)" and "large (>3 QALYs)", whereas the levels in the snowball and random surveys were "small", "medium" and "large".

As can be seen in Table 6.1, 'need' is the most important criterion for all samples and 'individual benefit' is the second most important criterion.

'Age' is the third most important criterion for the random and snowball samples, and fourth most important for the health services researchers. 'Lifestyle' is the fourth most important criterion for the random sample, fifth most important for the snowball survey and the least important criterion for the health services researchers' sample.

'Societal benefit' is the fifth most important criterion for the random sample, fourth for the snowball sample and third for the health services researchers' sample. 'No alternative

treatment' is the least important criterion for the random sample and the snowball sample, and second least important criterion for the health economist's sample.

For the three samples, the relative importance of the top two criteria ('need' and 'individual benefit') is almost twice that of any of the other four criteria. The largest percentage difference between 'societal benefit', 'age', 'no alternative treatment' and 'lifestyle' is 3.7% for the random sample and 3.5% for the snowball sample. This reveals that although the four criteria are still important to the respondents, there is little difference between them. The largest percentage difference between the weights for 'societal benefit', 'age' and 'no alternative treatment' for the health services researchers' sample is 2.7%. However, the weight on 'lifestyle' is much lower than the other samples at 5.8%.

As discussed in Chapter 4, it is somewhat controversial to use the criteria of 'age' and 'lifestyle' in a prioritisation process. However, respondents from the random sample placed more importance on these characteristics (14.2% and 12.8%) than on 'societal benefit' and 'no alternative treatment' (12.1% and 10.5%).

The criteria weights in Table 6.1 represent the relative importance of a criterion at its highest level (e.g. *large* 'individual benefit', *large* 'societal benefit'). Therefore the ratio of any two criteria weights represents the willingness of respondents to give up the maximum amount of one criterion (i.e. the highest level of the criterion) in order to achieve more of another. Because each criterion has a range of levels the ratio of any two criteria weights can be interpreted as an estimate of the *average* marginal rate of substitution (MRS) between the criteria (Amaya-Amaya et al. 2008). For example, the MRS between 'need' and 'age' is two (i.e. 28.4/14.2). This means that on average respondents consider 'need' to be twice as important as 'age'.

The next step is to use the average criteria weights to rank health treatments.

6.3 Ranking health treatments using criteria weights

In this section the vignettes from the focus group meetings and the criteria weights from the random sample are used to illustrate how health treatments can be scored and ranked using criteria weights.

First, the vignettes are entered into 1000Minds software and categorised according to the criteria on which they will be ranked. Figure 6.1 displays the 14 categorised vignettes (ignore the rank and total score for now). For example, to categorise a hip replacement, the appropriate level on each criterion is selected: 'this is the only treatment available', '15-64 years', 'small' (benefit to society) and so on.

As the vignettes are being categorised for illustrative purposes only, I categorised the 14 vignettes using the information supplied from health experts. This step is important as the categorisations may affect the overall ranking of treatments (which is explained shortly) and therefore in practice, ideally a health professional or someone who has expertise and knowledge in the treatment areas should categorise the treatments.

The 14 vignettes are categorised in relation to each. For example, a hip replacement is categorised as having a *small* 'benefit to others' when arguably this could have a *large* benefit, particularly if the person is the main income earner. However, compared to methadone treatment for instance, which is known to substantially reduce crime and the spread of hepatitis (Sheerin et al. 2004), a hip replacement could be considered as having a comparatively *small* 'benefit to others' compared to methadone which has a *large* 'benefit to others'.

The vignettes could also be categorised using more levels. For example, 'benefit to others' could be categorised on three levels – small, *medium* and large – by either including an additional level in the decision survey or extrapolating the criteria weights. Finally, when categorising the vignettes, an 'average' patient group is considered. For example, on *average*, patients receiving antiretroviral drugs for HIV receive a *large* individual benefit but this may not be the case for *all* patients.

Figure 6.1: Categorising the vignettes according to the criteria and levels

	CRITERIA							
PATIENT click to open	Treatment options for this patient	Age of patient	Benefit to others (*eg* family or society)	Patient's health before treatment	Benefit to patient (*ie* length and/or quality of life)	Illness or injury caused mainly by lifestyle choices	RANK	TOTAL SCORE
Hip replacements	this is the ONLY treatment available	15-64 years	small	poor (but not immediately life threatening)	large	no	1st	66.9 %
Dialysis for renal disease	this is the ONLY treatment available	65+ years	small	will die soon without treatment	large	yes	2nd	61.0 %
Imatinib mesylate for chronic myeloid leukaemia	this is the best treatment (there are less effective alternatives)	15-64 years	small	will die soon without treatment	medium	no	3rd	60.7 %
Antiretroviral drugs for HIV	this is the ONLY treatment available	15-64 years	small	poor (but not immediately life threatening)	large	yes	4th	54.1 %
Abatacept for rheumatoid arthritis	this is the best treatment (there are less effective alternatives)	15-64 years	small	poor (but not immediately life threatening)	medium	no	5th	46.5 %
Growth hormone treatment for Prader-Willi Syndrome	this is the ONLY treatment available	0-14 years	small	fair (neither good nor bad)	small	no	6th	44.2 %
Hand sanitiser use in primary schools	this is the best treatment (there are less effective alternatives)	0-14 years	large	relatively good (though treatment is still beneficial)	small	no	7th	39.0 %
Service for postnatal depression	this is the best treatment (there are less effective alternatives)	15-64 years	large	fair (neither good nor bad)	small	no	8th	39.0 %
Methadone	this is the best treatment (there are less effective alternatives)	15-64 years	large	poor (but not immediately life threatening)	small	yes	9th	33.6 %
PET Scan	this is the best treatment (there are less effective alternatives)	15-64 years	small	fair (neither good nor bad)	small	no	10th	26.9 %
IVF Treatment	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	20.2 %
Oral drugs for erectile dysfunction	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	20.2 %
Vaccine for cervical cancer	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	20.2 %
Statins for cardiovascular disease	this is the best treatment (there are less effective alternatives)	15-64 years	small	fair (neither good nor bad)	small	yes	14th	14.1 %

When each vignette is categorised according to the criteria levels, the associated weights from the random sample are automatically assigned to the levels which enables an overall score to be calculated for each vignette. For example, in Figure 6.2 an overall score for 'hip replacements' is obtained by summing the relevant weights for each criterion.

Figure 6.2: Scoring a hip replacement using the random sample weights

Treatment options for this patient		Score
this is the best treatment (there are less effective alternatives)	0.0 %	10.6
this is the **ONLY** treatment available	10.6 %	
Age of patient		
65+ years	0.0 %	7.4
15-64 years	7.4 %	
0-14 years	14.2%	
Benefit to others (*eg* family or society)		
small	0.0 %	0
large	12.1 %	
Patient's health before treatment		
relatively good (though treatment is still beneficial)	0.0 %	14.1
fair (neither good nor bad)	6.7%	
poor (but not immediately life threatening)	14.1%	
will die soon without treatment	28.4 %	
Benefit to patient (*ie* length and/or quality of life)		
small	0.0 %	22
medium	12.2 %	
large	22.0 %	
Illness or injury caused mainly by lifestyle choices		
yes	0.0 %	12.8
no	12.8 %	
	Total Score:	66.9

As there are no alternative treatments for a hip replacement, a hip replacement scores 10.6% on the first criterion; the average age of a patient requiring a hip replacement is 50, so a hip replacement scores 7.4% on the second criterion and so on. Summing the criteria weights gives hip replacements a total score of 66.9%.

Once all vignettes have been categorised and scored, the vignettes can be ranked based on their overall scores. Referring back to Figure 6.1, it can be seen that 'hip replacements' has the highest overall score. Conversely, 'statins for cardiovascular disease' is ranked last with a total score of 14.1%. (On average, this treatment provides only a small benefit to the patient and the patient's health before treatment is not poor which is in contrast to the high need and high benefit associated with hip replacements.)

The ranking of the vignettes is sensitive to the criteria levels changing. For example, if the benefit to a patient of receiving dialysis was reduced from 'large' to 'medium', dialysis will score 12.2% instead of 22% on this criterion, lowering its rank from second to fourth place. The sensitivity to changes in categorisations can be ameliorated by using interpolated levels. For example, with 'benefit to patient', two new levels could be approximately interpolated – 'small and medium 6.1%' (which is the average of the 'small' and 'medium' weights) and 'medium and large 17.3%' (which is the average of the 'medium' and 'large' weights).

Of course, these rankings relate to the *benefits* of the treatments only and do not take into account cost or any additional considerations such as 'strength of medical evidence'. Once these additional factors are considered, the overall ranking might change.[145]

6.3.1 Using vignette rankings as a 'gold standard'

As discussed in Chapter 5, one way to check the validity of a 1000Minds decision survey is to use *directly* ranked vignettes (i.e. vignettes ranked by group agreement) as a 'gold standard' against which *indirectly* ranked vignettes (i.e. vignettes ranked according to their overall score) can be compared. This approach is particularly relevant when the group directly ranking the vignettes are health experts or are very knowledgeable in the subject area and/or they have experience in prioritising health services. As the focus groups in this thesis were *not* established based on respondents' knowledgability or experience, and as some of the terminology used in the focus groups was different to the terminology used in the decision survey (e.g. 'number of patients' was included in the health vignettes but was not included in the decision survey), it is not feasible to use the focus groups rankings as a 'gold standard'. However, to illustrate how this can be done, the consensus rankings from the focus groups and the vignette rankings from the three sample groups will be used (shown in Table 6.2).

[145] This is discussed in Chapter 9.

Table 6.2: Rankings of health vignettes by the focus groups and the random sample

Health vignette	Nurses *N*=5	Non-medical health workers *N*=4	Retirees *N*=7	Public health *N*=6	GP Practice *N*=5	Maori health provider *N*=13	Random sample *N*=322	Snowball sample *N*=275	Health services researchers' sample *N*=12
Statins for patients at high risk of cardiovascular disease	1	1	2	1	1	2	14	11	11
Hip replacements	11	2	3	4	2	4	1	1	2
Service for postnatal depression	3	7	4	2	6	1	8	7	8
Methadone for opioid addiction	5	6	10	7	4	5	9	9	6
Dialysis for end-stage renal disease	7	10	7	8	7	7	2	2	1
Vaccine for preventing cervical cancer	13	5	11	5	3	3	11	10	12
Positron emission tomography (PET Scan)	12	4	1	10	11	6	10	11	10
IVF treatment	4	8	8	6	8	9	11	11	12
Imatinib mesylate for chronic myeloid leukaemia	2	11	6	13	10	11	3	3	4
Antiretroviral drugs for HIV	10	9	9	9	5	10	4	4	3
Abatacept for last-line treatment of rheumatoid arthritis	6	12	5	12	9	8	5	5	5
Hand sanitiser use in primary schools	9	3	14	3	12	12	7	8	9
Growth hormone for Prader-Willi Syndrome	8	13	13	14	13	13	6	6	7
Oral drugs for erectile dysfunction	14	14	12	11	14	14	11	11	12

To measure the level of agreement across three or more different groups, of the vignette rankings, Kendall's coefficient of concordance (Kendall's W) is used (Norusis 2008). Kendall's W ranges from 0, indicating no agreement between the groups, to 1, indicating complete agreement between the groups. As discussed in Chapter 3, Kendall's W across the six focus groups is 0.553 (p=0.000), indicating moderate agreement. However, when Kendall's W is calculated across seven groups (the six focus groups and one of the sample groups) there is only weak to moderate agreement. For instance, Kendall's W = 0.4 (p=0.001) across the six focus groups and the random sample, Kendall's W = 0.359 (p=0.002) across the six focus groups and the snowball sample and Kendall's' W = 0.330 (p=0.004) across the six focus groups and the health services researchers' sample.

As well as comparing the level of agreement between the ranked vignettes from the six focus groups and the indirectly ranked vignettes from each of the sample groups, the rankings from

individual focus groups were also compared to the rankings obtained from a 'comparable' sub-group within the random sample. For example, the rankings from the 'retirees' focus group were compared to the indirect rankings of random sample respondents who are over 65 years. Similarly, the rankings from the Maori health provider focus group were compared to the indirect rankings of random sample respondents who identify as Maori. Spearman's rank-order correlation was used to measure the level of association between *two* groups. All correlations are small and insignificant.

A possible reason for the lower concordance between seven groups (the six focus groups and a sample group), and for the low correlations between individual focus groups and similar sub-groups from the random sample is that respondents did not consider exactly the same criteria. As mentioned earlier, the vignettes in the focus groups were described at an overall treatment level and included the number of patients involved. In many focus groups the number of patients treated became the main priority when ranking the treatments.[146] In contrast, respondents in the decision survey were asked to choose between two hypothetical patients. The number of patients affected by a treatment was not a consideration. Therefore the participants in the focus groups and respondents from the random sample were not 'trading' exactly the same criteria. In addition, ranking vignettes is not easy (for an individual or a group) as multiple conflicting criteria have to be considered at the same time. In contrast, respondents completing the decision survey considered only two criteria at a time.

As mentioned previously the focus groups were not established based on knowledge and experience and therefore their rankings should not be considered as a 'gold standard'. An example of how group rankings have been used as a gold standard is discussed in Hansen & Ombler (2008). In 2004 a group of cardiologists and cardiac surgeons from the New Zealand Region of the Cardiac Society of Australia & New Zealand, supported by the Ministry of Health, created a value model using 1000Minds software to prioritise patients for coronary artery bypass surgery (Hansen & Ombler). The authors found that the overall ranking of a set of patient case descriptions resulting from the value model was highly correlated with the clinicians' initial ranking based solely on their judgements. However, it should be noted that even when a high Kendall's W is achieved, this merely indicates that there is a high level of agreement between the groups. It does not imply that the ranking is correct.

[146] For example, in the description for "statins for patients at high risk of cardiovascular disease" it was stated that 220,000 people at risk of cardiovascular disease will receive statins for the rest of their lives potentially preventing 66,000 heart attacks or coronary deaths. These high numbers became the focus for many group members, resulting in four of the six groups ranking this vignette first.

6.4 Preferences from the literature

Only one other study has used 1000Minds software and the PAPRIKA method to elicit weights for criteria relating to health care. Golan et al. (2011) surveyed a convenience sample of 74 respondents to establish the relative importance of five criteria. The criteria are as follows, with the associated weights in parentheses: lives saved (0.343), life-prolongation benefits (0.243), quality of life gains (0.217), alternative treatments (0.107) and other important social/ethical benefits (0.087). The criteria configurations used in the Golan et al. study differ from the criteria used in this thesis but comparisons are still possible. As can be seen from the above criteria weights, the most important criteria are 'lives saved', 'life-prolongation benefits' and 'quality of life gains'. These criteria can be compared to 'need' and 'individual benefit' in this thesis, which were deemed by respondents to be the most important criteria. Similarly, 'alternative treatments' received the second lowest weight in the Golan et al. study and the lowest weight in this thesis. It is difficult to compare 'other important social/ethical benefits' as this criterion is essentially three criteria in this thesis – 'benefit to others', 'age' and 'lifestyle choices'. However, the combined average weight for the three criteria is much higher than the weight for 'other important social/ethical benefits' in the Golan et al. study.

It is not possible to compare the criteria weights from this thesis with criteria weights from other studies because of differences in methodologies, criteria and health settings. However, a review of the literature suggests that 'health status' (need) and 'patient benefit' tend to be the most important factors when prioritising health care (Mullen 2004, Menon & Stafinski 2008, Sabik & Lie 2008, Diederich et al. 2011). Although some studies mention social and ethical factors such as age and lifestyle, most studies do not include these factors explicitly in a prioritisation process.

The results of this survey indicate, in a clear and transparent way, that for a randomly-selected sample of the New Zealand public, 'need' and 'individual benefit' are the two most important considerations in prioritising health care and that factors such as 'no alternative treatment', 'benefit to society', 'age' and 'lifestyle choice' are also important albeit to a lesser extent.

Many respondents took the opportunity to comment when answering the decision survey. The comments provide some context as to how respondents made their decisions. The comments and feedback on the survey are discussed in the next section.

6.5 Respondent feedback

Survey respondents were invited to make comments throughout the survey. They could write comments in an allocated space beneath each trade-off question and/or they could comment on the survey overall or health care prioritisation in general at the end of the survey. More than 150 respondents from the random sample made comments. Some respondents phoned me or wrote letters or emails, 50 respondents commented in the 'trade-off' part of the survey and 95 respondents commented at the end of the survey. Most comments related to the criteria or were explanations relating to the trade-off question being answered. Out of 322 respondents in the random sample, 281 respondents requested a copy of the results. A summary of the results sent to respondents by email is in Appendix 6.1.

Comments were also received from respondents in the pilot and snowball samples. As discussed in Chapter 5, the snowball sample is not demographically representative of the New Zealand public (e.g. only 4% of the snowball sample respondents were over 65 years compared with 17.3% of the population, 0.4% of the snowball respondents had no qualifications compared with 18.7% of the population and 31.3% of the snowball sample had a household income over $100,000 (not taking into account the respondents who chose not to give their income) compared with 16.8% of the population). Because the criteria weights from only the random sample will be analysed (in Chapters 7 and 8), only the comments from the random sample respondents will be discussed.

Qualitative data need to be analysed in a systematic and rigorous manner to ensure that an accurate interpretation of the data is presented. Computer-aided Qualitative Data Analysis software (CAQDAS) is available to help with transcribing, coding, interpreting and extracting main concepts from qualitative data. Other methods of qualitative data analysis include phenomenology (van Manen 1990), grounded theory (Glaser & Strauss 1967) and content analysis (Krippendorf 2004). These methods use interviews and/or observations to record and code information for the purpose of understanding a particular phenomenon, establishing a new theory or testing an existing one.

In this thesis the qualitative data were not collected by means of interviews or observations and therefore the above methods are not suitable for analysing the qualitative data. Instead the technique used in this thesis is similar to the process suggested by Warden & Wong (2007) for analysing qualitative data: data collection, note taking, coding (highlighting main ideas),

identifying themes, providing illustrative quotes, recognising saturation (when collecting more data provides no new information), memoing (identifying potential themes or relationships) and sorting (providing a narrative by compiling and arranging themes).

The respondents' comments, grouped into main themes, are outlined and discussed in the following sections. Most comments relate to the six criteria. Other comments relate to the survey design, how respondents made their decisions or on health care prioritisation in general. A full catalogue of respondents' comments is in Appendix 6.2.

6.5.1 Age

> "Age SHOULD NOT come into the choices of what is right and what is wrong for us all."
>
> "The young ones should be looked after first as (hopefully) they have a longer future."
>
> "We need our 65+ year olds to help teach the younger generation."

Three respondents commented that the age ranges of 0-18 years, 35-64 years and 65 years and above were too broad and that narrower bands would be better. For example, one respondent suggested that age groups should be categorised as follows: new-born, infants, young children, young adults and adults. Four respondents felt that the 65 years and over category was too 'young' to be considered the oldest category and that another category, perhaps over 75-80 years, should be included.

Three respondents commented that older people are still productive and, with the retirement age likely to rise, more importance should be placed on the productive (and usually older) members of society. Four respondents commented that older people have contributed to society for many years and therefore they should receive priority (or at the very least, not miss out) when it comes to health care. Four other respondents felt that young people still have "time on their side" whereas older people do not and therefore they should receive priority. Conversely, 14 respondents felt that younger people should always receive priority as older people have already had a 'fair innings' and that younger people have more years ahead of them to contribute to society. Five respondents felt age should not be considered at all.

As discussed in Chapter 4, all criteria levels, including age levels, needed to be entered into the 1000Minds survey in order of importance. It is straightforward to rank in order of importance 'individual benefit' (i.e. small, medium and large) but it is not so easy to rank age in order of importance. However, it is *generally* accepted that preference is given to younger people when it comes to health care prioritisation (Williams 1997), though it gets difficult to order the age categories in terms of value or importance when there are many age categories to consider. It is therefore less controversial to use three age categories that broadly encompass three segments of society: retired people, working-aged people, and babies and children.[147] This is certainly an area in which further research in New Zealand would be worthwhile.

6.5.2 Alternative treatments

> "It is the only option for this person."

Only one respondent commented specifically on this criterion.

6.5.3 Benefit to individual

> "A small benefit can feel enormous to an unwell patient."

> "Having been born with a congenital deformity I am strongly in favour of making sure that only the strongest in our society survive. Being different in this life is no fun at all."

> "I also don't agree with surgery for SOME congenital/hereditary illnesses where in the natural order of things, they would die."

Six respondents felt that the benefit to an individual of receiving treatment was important, even if that benefit was small. Two respondents commented that benefit to an individual was important so long as the treatment left them with a reasonable quality of life. Five respondents felt that too much emphasis is placed on extending life at the expense of quality of life, particularly when co-morbidities exist.

[147] As mentioned in Chapter 4, these three age groups were used by the New Zealand Treasury in their report on Population Aging and Government Health Expenditure (New Zealand Treasury 2005).

6.5.4 Benefit to society

> "It is difficult to know just when in a person's life-path they can be of a benefit to society; some contributions come later."

The comments relating to this criterion revealed that some respondents interpreted 'benefit to society' in a way different from what was intended. Benefit to society relates to the flow-on effects of patients receiving treatment. For example, the flow-on effect of children being immunised is that they will not spread illnesses to other non-immunised children (and parents). Comments received from eight respondents indicated that they interpreted 'benefit to society' to mean that one of the two hypothetical patients provided a greater benefit to society than the other, not in terms of flow-on effects but more in terms of their age or occupation for instance.

6.5.5 Lifestyle

> "Save a life without judgement."

> "A patient who has made a conscious choice to be unhealthy must bear the consequences and be further down the line."

Respondents commented on 'illness or injury caused mainly by lifestyle choice' more than any other criteria. Four respondents commented on lifestyle when they were trading-off two criteria that did *not* include 'lifestyle'.

This criterion polarised many respondents. Four respondents felt that everyone should be treated equally regardless of why they became ill. Three respondents commented that for some people their lifestyle may not be a 'choice' but be imposed on them because of lack of money for instance. Similarly, five respondents commented that children do not make lifestyle choices, that their lifestyle is dictated by their parents and therefore young people should not be disadvantaged if their illness has been caused by their lifestyle. Nine respondents felt that poor lifestyle choices should not disadvantage patients but that treatment should be given on the proviso that patients agree to change their current lifestyle (for example, to give up smoking).

Twenty one respondents believed that priority should be given to the patient whose illness was *not* caused by their lifestyle choices. However, for 15 of those respondents it depended on what constituted a poor 'lifestyle choice'. For example, if patients were smokers or heavy

drinkers, then they should not receive priority. On the other hand, if patients had illnesses or injury as a result of 'healthy lifestyle choices' (e.g. a marathon runner requiring a hip replacement) then they should not be penalised. Although respondents have different views of what is considered to be a poor 'lifestyle choice', it is an important factor in priority-setting.

6.5.6 Need

> "I would assume anyone of any age, whatever the long term prognosis, should be treated first if they would die without it."

Ten respondents commented that people in the greatest need, particularly those facing death without treatment, should be treated first. Three respondents commented that it is better to treat patients before they become very ill and require greater amounts of treatment. Similarly, three respondents believed that the focus should be on preventative measures.

6.5.7 General comments

> "Difficult decisions to make – economics or ethics or social status?"

> "The stance I took is based on all people being equal and therefore entitled to health care as a right. This is what taxes are for."

> "Good management should bring with it efficiency and productivity."

Eight respondents acknowledged that prioritisation of health service is necessary, but that it is a very difficult area, often fraught with emotion, involving decisions that they would never wish to make.

Twenty seven comments related specifically to the trade-off questions. Most were justifications of why a respondent chose one patient over the other. Four other comments related to how difficult it was to make the decision. Three respondents suggested that more information was needed to make a decision. Eight respondents made comments relating to their personal circumstances. Eight other respondents mentioned the absence of information concerning the cost and effectiveness of treatment in the survey.

Four respondents commented that everyone deserves to be treated equally and that, for example, “fairness is the critical cement that holds a civil society” together. Two respondents believed that “health care is a right” and that is why we should all be treated equally.

General comments relating to New Zealand’s health system included suggestions that its better management will promote efficiency and productivity.

It was pleasing that more than 150 respondents commented on the trade-off questions and health care prioritisation in general. Not only does this demonstrate that respondents were interested in the topic and understood the survey, but also the comments provide some insight as to how trade-off decisions were made.

6.6 Conclusion

In this chapter the main findings of the decision survey were discussed. For the three samples (random, snowball and health services researchers) ‘need’, and ‘individual benefit’ were revealed as being the two most important criteria. ‘Age’, ‘lifestyle’, ‘societal benefit’ and ‘only available treatment’ were also considered important but to a lesser degree.

The many comments received by respondents have contributed to a rich data set providing insights into the decision-making processes of respondents and their thoughts on health prioritisation in general.

Appendix 6.1: Email of results sent to random sample respondents

HEALTH CARE PRIORITISATION IN NEW ZEALAND

In December last year you completed on on-line decision survey relating to health care prioritisation. You indicated that you would like a summary of the results. A summary of the preliminary results follow.

Over 3200 letters were sent to people randomly selected from the electoral roll. 322 people completed the survey. Although this is not a high response rate we were still able to analyse the survey responses. Your participation has made this possible, so thank you.

As you may remember, you were asked to trade-off certain characteristics that could be considered when prioritising health care treatments. These were: age of the patient, benefit to the patient, benefit to society, whether an alternative treatment was available, health status of the patient and lifestyle. From the survey responses we have calculated "weights" for each characteristic, that is, how important each characteristic is in relation to each other. This is what we discovered:

Weights for prioritising health care treatments:

Patient's health before treatment (health status)	28.4%
Benefit to patient (length and/or quality of life)	22.0%
Age of patient	14.2%
Illness or injury NOT caused by lifestyle choices	12.8%
Benefit to others (eg family or society)	12.1%
Only treatment option available for this patient	10.6%

1. The survey participants, on average, think that a person's health status (how sick they are before treatment) is the most important characteristic, followed by the benefit that person will receive from treatment. Whether an alternative treatment is available, is the least important characteristic.

2. The percentages represent how important each characteristic is in relation to each other. For example a person's health status (28.4%) is twice as important as the age of a patient (14.2%). The benefit to a patient in terms of length and/or quality of life (22%) is almost twice as important as benefit to others (12.1%).

3. It may be difficult to use the characteristics of age and lifestyle in a prioritisation process because of discrimination. However, the results show that on average, participants do place an importance on these characteristics (14.2% and 12.8%) and consider them to be slightly more important than benefit to others and treatment options (12.1% and 10.6%).

4. Where to next? The benefits of health treatments (using these weights) together with cost and other considerations such as treatment effectiveness, can be used to compare and prioritise health services.

Thank you for taking part in this survey. This research is very important and a basis for more in-depth research. Thank you also to all those people who took the time to write additional comments.

Trudy Sullivan

Appendix 6.2: Comments from random sample respondents

Age
▪ Age SHOULD NOT come into the choices of what is right and what is wrong for us all - we ALL deserve the very best that the health system can offer. ▪ To serve the greatest good they must be treated equally (trading off age and benefit to society). ▪ The young 0-15 are our future and illness could affect their psyche for a normal healthy life in society - hence me favouring them. 15-64 are taxpayers hopefully and deserve consideration. 64 upwards hopefully have paid taxes and deserve something from the health system. ▪ I believe each patient should be treated in order of merit regardless of age. ▪ I think kids should take priority. ▪ In my opinion the young ones should be looked after first as {hopefully} they have a longer future. ▪ Younger people should be considered first as hopefully they have years to contribute to the good of NZ. ▪ Basically children 15 years and under haven't lived a length of life like an elderly person - 65 years and older so if they need treatment right away because their life/health is at risk in some way, they need to be given treatment first to have the chance to keep living happily. ▪ Very difficult but I think we must support our future generations. ▪ Save the young. Too many people afraid to die when their time is up. ▪ There is already a greater societal investment in the 'working' person (why favouring younger person) ▪ I would always treat the younger one first to ensure no long term problem. ▪ I selected this age group (youngest) due to that fact that more and more kids die because they are not fully immunised to the viruses out there. They also stand to be at a higher risk of infection. ▪ At the age of 65 we should be able to support ourselves better and afford to pay for the treatment if needed. The younger patient should be treated first. ▪ I think that the older the patient the less likely I would be to choose this response. ▪ I feel that older people have just as much right to a decent health service, even though they may be less useful to society, but remember they too worked and paid taxes when they were younger. ▪ As the retirement age is going to rise and elders are going to stay self-managing for longer it is sensible to ensure their capacity to do so. ▪ I am 64 years of age and I choose this one for the simple reason we can't retire yet - we are the big tax payers and we are necessary to prop up a fast failing revenue that could hit the wall soon. ▪ We need out 65+ year olds to help teach the younger generation. Without our 65+ year old we lose sight of our past, and never learn from mistakes that had occurred. ▪ (The younger person) will have time to make other choices. (The older person) no choice but to be given treatment. ▪ (Oldest group) In that age range the benefit is worthwhile. ▪ The other patient (youngest) has time on side. ▪ Whilst the younger patient needs the treatment he /she is able to be looked after by mum, dad etc., while the older patient may need special carers because of responsibilities etc ▪ The vast age ranges make the divides less significant than the range within a category. ▪ It is quite difficult to grade children (for eg new-born, infants, young children, 0-5years) with young-adults or adults (5 years and above) because the babies could deteriorate in health condition due to infection etc very fast as compared to young adults and adults.

Hence, treatment of children (below 5) should be treated separately than the rest.

- My choices were inconsistent- because I realize (at 51) that 65 years of age isn't very old! However, 2 years ago my then 16 year old suffered a stroke. He has made a 100% recovery thanks to the public health system. I consider this money well spent; he has a lifetime to repay this with his taxes.
- Again I think that there is a cut off for my choosing the "right" answer. (My answer would be different if the age was 75). I think age becomes more of a factor but most people of 65 can still lead productive lives.
- My decision may be changed depending on how many + years have passed, particularly if the patient is over 75 years of age.
- If you had focussed on those over 80 rather than over 65 it may have made a difference as I would support making them comfortable, but not to spend too many $ on medical care than may only give them less than 1 yr to live.
- I found the cut-off age of 65+ not old enough to make the decision that they've had a good innings and shouldn't receive as much urgent treatment, as a younger person.
- Although 81 I am very concerned about what I see as a bias in our public health system towards caring for the old at the expense of the very young who do not have votes. .. I am grateful for the treatments I've received over the past 14 months, especially the eyelid operations but the thyroid treatments do raise a question. The cancer there could well move slowly since I am 81. Should I have been put at some risk by being required to wait while some child received some alternative treatment? I guess this is where your survey comes in. How serious should a child's illness need to be to take precedence? Had an illness been at all serious, or with long term adverse consequences I would certainly have been prepared to give up my place in the queue if asked. I guess - as often happens in life - the situation is more complicated because I am my husband's carer - he has Alzheimer's and is reaching a tricky stage. If, as does not happen to be the case, the State had to pay for his care, keeping me alive would save the health system about $42,000 annually. That should pay for quite a lot of treatments for young children.
- I am not sure how you prioritise people but the biggest population in NZ is like myself on a very low income and not able to afford to go and see a doctor. And some of the doctors in this country are imports and really do not care about their patients. I just had an experience with a locum from Germany: I went to see her about a spot on my head and she treated me like an infant I am 62yrs of age and a surviving cancer person. So the health really needs to be looked at young, medium and old. We all as human beings deserve to be treated as such for survival. I also have a newborn grandson{5mths old} who has a heart problem so all the help my family could get I would be humbly grateful.
- Thinking of my family members there would be different answers depending on the age range. 15 - 64 is quite a wide range and obviously the older someone is the variables would come in to it. eg: my mother of 90 would prefer non-intervention for herself in favour of someone with a life ahead of them but that's all a different question - fascinating - I wish you well in your research.

Benefit to Individual

- Depends on co-morbidities in the older individual and what benefit is going to be achieved with treatment - i.e. what quality of life are you going to give them by treatment. Extending life may not give them quality of life. For the younger patient again it depends on the benefit of treatment.
- It is important to see the patient as an individual worth saving.
- A small benefit can feel enormous to an unwell person.
- This one (high benefit to patient) because the long term and life quality is greater.

- It's the patient not society that needs the treatment.
- When viewing the questions, my main concern was to the patient, not the close family or relatives.
- Better chance of this patient having a good life (large benefit/fair health > medium benefit/poor health)
- Virtually everyone carries some 'impairment' and small improvement for an individual not seen as a priority.
- If the patient has a reasonable chance of returning to a functioning member of family and society.
- I have strong views about prolonging lives that are not going to be good for the patient or their family. I also don't agree with surgery for SOME congenital/hereditary illnesses where in the natural order of things, they would die. My reason for this is, if the person lives to have a family, the same condition could be not naturally bred out, and be passed onto the next generation. A harsh opinion I know. Every case is unique and should be assessed individually (ideally).
- Having been born with a congenital deformity I am strongly in favour of making sure that only the strongest in our society survive. Being different in this life is no fun at all!! Health cuts to supplies, services and the stupid way prescriptions are administered is both time wasting and frustrating for those of us that must take medication to survive. I would rather have paid an extra $5 per week towards my condition than thousands to Southern Cross over the years which has done nothing for me.

Benefit to Society

- (Large benefit to society) Prospects better for all concerned.
- Societal benefit should come before individual gain / benefit.
- It is difficult to know just when in a person's lifepath they can be a benefit to society, some contributions come later.
- The "benefit" gives me a problem. Who decides on how much benefit to others they represent? Is a fireman of more value/less than a teacher?
- Who decides the "benefit to others"? (Not comfortable with this.)
- I feel the younger - we should all be equal whether some perceived benefit to society exists - the view of whom is important could be subjective. (when trading off age and benefit to society)
- It would be difficult to state who is important to society as I can see if that were a criterion many people deserving or not could jump the queue. Importance to society might be in the eye of the beholder.
- I felt very uncomfortable with the "benefit to others" factor - who would judge this? How would this be judged? I did not feel I could choose between these options, probably because I believe every life is valuable.
- I found the judgemental assumption around benefit to family society and others a slightly loaded premise. Is a sick mother of more or less value to society than a future chemist who might discover the cure for cancer? Your question tends to combine nurturing value with worth to the wider community. I would imagine the clinician would hardly be able consider the latter when presented with any patient.

Lifestyle
▪ Not sure if a 14yr old could have injuries as a life style choice. The lifestyle of a 14 yr old is that dictated by the parents - not a choice of the child. ▪ 14 year olds don't make lifestyle choices, they are made for them. ▪ What lifestyle choices is a kid 14 or younger likely to make? ▪ An interesting one - lifestyle choices for under 15 year old - could be dangerous sport or eating wrong diet. However I think should still take priority (age vs lifestyle) ▪ After the age of 21yr I would take into consideration lifestyle choices and compliance in health issues if the person has had the benefit of being exposed to and making these choices. ▪ Save a life without judgement (need vs lifestyle). ▪ If the patient needs help this should be done, regardless of his lifestyle. ▪ We all make lifestyle choices good or bad and no-one is perfect also person could also have a hazardous occupation e.g. working with asbestos ▪ I am not prepared to penalise an individual for bad decisions. It could involve an operation for lung cancer or the treatment of emphysema for example. I would not want to condemn these people, especially as the other patient has treatments available. ▪ Every human being has value. A person's lifestyle may have affected their health but there are usually other factors that influence lifestyle that are involuntary. ▪ Sometimes lifestyle choice is due to lack of money for health living. ▪ Also people may be misinformed about the health effects of certain lifestyle choices, when that choice becomes life threatening who is to blame? Provided necessary changes to lifestyle are agreed to. ▪ Lifestyle choices are the responsibility of the individual and if they intend not to change this, it is better to treat the other patient first. ▪ Here I would expect the individual in the left hand box (poor lifestyle) to be asked to change their behaviour, and give preference to the individual who has behaved sensibly. ▪ They need to make a commitment to change lifestyle. ▪ With change of lifestyle a requirement. ▪ This patient, if change in lifestyle is initiated should have a better life. ▪ May have chance to change lifestyle (no alternative treatment vs benefit to patient) ▪ I am most definitely in favour of making people change their lifestyle to gain improved health. ▪ I also have strong views on not providing health care for lifestyle choices that endanger lives. People who smoke or are obese should not be given treatment until they get rid of both. ▪ The patient has caused less of the problem and all people should be equal and not judged on a perceived benefit to society. (good lifestyle/small social benefit > bad lifestyle/large benefit to society). ▪ A patient who has made a conscious choice to be unhealthy must bear the consequences and be further down the line. ▪ "A no brainer" (to choose the one with the good lifestyle). ▪ Individual responsibility. ▪ Depends again on what the required treatment is - for example if a drunk driver comes in and needs treatment but the person he/she has caused injury too, should be treated first. ▪ Health care prioritisation - a bit issue from here on - good life style choice should be a prerequisite. ▪ Lifestyle choices--too large a field, should be broken down. ▪ Everyone has to eat food; it can be abused but shouldn't be in the same category as drugs, smoking, booze, accidents. ▪ I think that in cases where there are specific causes ie direct from smoking then other

patients should be treated first.
- Particularly if poor prior health is self-assisted eg by smoking, obesity, alcoholism (age vs need).
- It's hard to make a determination on such limited information - the "left" patient may be suffering from self-imposed health issues, which should not give them an advantage over another. (health status vs benefit to patient)
- This is tricky considering if an injury is caused during a healthy choice eg: running, then that would differ much from someone that smokes and their lungs are damaged.
- Assuming it was an unhealthy choice of lifestyle???
- Poor health due to lifestyle choices? I thought this a bit vague, were they negative lifestyle choices like smoking? Or positive lifestyle choices like marathon that can screw your knees up.
  Nutritional education: Change to exercise regime if repeated injuries caused by contact sport.
- Some questions that refer to lifestyle choice were conflicting, eg if the lifestyle choice was smoking then I would likely answer that the other patient should receive treatment, if the lifestyle choice was sports then I would reconsider.
- It wasn't clear who made the judgement calls -- on what, for example, constituted a health-threatening lifestyle choice by the patient, and whether he/she was aware of it.
- Depends on what their issue in society was – crimes (when trading off individual benefit and another, ie not asked about lifestyle).
- The questions including lifestyle options were too vague (undefined) as there is a big range of "life style" choices which can affect your health. Because of this I'm not too happy that a life style choice could be relevant as to whether you get treatment or not.
- As noted in the survey, some of the questions related to the injury/illness being of lifestyle choice needed to be qualified more I feel. That is, I would have answered differently if were someone becoming injured making a good lifestyle choice like running for health, rather than someone making a poor lifestyle choice such as smoking.
- In my opinion health or injuries requiring treatment as a result of lifestyle choices should not take precedence over health or injuries resulting from unavoidable accidents or naturally occurring events. The cost of treatments resulting from road accidents where alcohol is the prime cause should be borne by those causing the accidents or their insurers. The resulting savings to the health system would enable more treatments being available other patients.
- On trade-off questions would have preferred to know more specifics about how person was responsible for their condition, e.g. smoking - lifestyle - drinking etc.
- If I had to choose between someone receiving treatment as a result of a sports injury or someone who has aggravated their condition by smoking then I would always choose the sports injury. That person was presumably concerned with body fitness whereas the smoker was not.

Need

- Always concerned with imminent death if treatment can avoid this outcome.
- The patient will die without treatment.
- I would assume anyone of any age, whatever the long term prognosis, should be treated first if they would die without it.
- If it is the final option then this patient MUST be treated first (trading off age and die soon).
- (Die soon) must take preference.
- The pattern of my choices so far seems to reflect the seriousness of the condition being

treated.
- The patient that has the worst health problem should be treated first.
- I think prioritisation must be given to ones in critical condition.
- More serious problems in the health of the patients, may have elicited a different response from me if those patients had a minor ailment!
- Would imagine that prioritisation would occur on seriousness of health issue rather than age
- Better to treat before gets really bad.
- I have never understood why we do not teach and practise preventative medicine more widely.
- Would like to see more emphasis on prevention rather than focus on critical cases even though I ranked these highly.
- So much money and time is wasted with waiting until people are really sick, before receiving treatment.
- The (sickest) patient may be at a stage where living longer is mere survival. If this is the case I would not want to prolong life. In contrast there is a clear benefit in treating the (not so sick) patient. So this is why I am choosing this box in contrast to my other selections by results, where ordinarily I tend to give great emphasis to saving life.
- We need to go back to fixing medical problems as they arise, not waste resources and lives by leaving treatment to the final stages.

General

- There are huge issues facing our provision of Healthcare and how we prioritise. Keeping people alive through medical intervention with no quality of life and at huge expense, particularly the very old, seems a waste of precious funds which could be better spent.
- I'd hate to be the one who has to make the ultimate decision – if there's only enough time, manpower, money or medicines for one person.
- I would hate to be the person who has to make these decisions. More Government funding for health could alleviate some of these problems but obviously sometimes a decision for "first in line" has to be made. When in an emergency and the midst of decision making those who are closest to me would be my top priority.
- Difficult decisions to make - economics or ethics or social status? There are so many other criteria that will help the professionals make the decisions about treatment.
- I think there will always be tough calls to make in health prioritisation. I think it is quite an emotive topic.
- I think that overall, priorities are managed well.
- Healthcare is a lot of conflicting tradeoffs. I think that putting resources into evidence-based healthcare and banning direct to consumer marketing would be a good start. Advertising pharmaceuticals on TV etc is a way to misallocate resources and reduce the overall health of New Zealanders. Admitting that dieting is almost always counterproductive would also improve people's health and allow better use of resources.
- It is very sad that Health Care has to be prioritised. Would rather see a compulsory private system than the one we have at present with many pensioners who have paid taxes throughout their life being left out of the system. The pressure being put on the countries A & E Depts is huge as this is the only avenue local GP's have to ensure these people get the care they need.
- I feel that making the decisions are easy if you have a strong set of guidelines to follow and if you are not involved with a person. I feel that specialists and doctors are not necessarily the right people to be making decisions and people can be swayed by having any personal involvement at all and definitely people can call in favours and influence

decisions by knowing the right person etc. It is my opinion that in our health system we need to develop a category of health professional that is a facilitator in that even at very basic level care particularly in communities people have access to a person who can sit them down and advise them of choices and varying health professionals that they could seek help from with regard to their particular problem. This person would have no alliance with pharmaceutical companies or mainstream western medical practice. They should have a good knowledge of western, eastern and alternative practices and they should not be taking any handouts or commission from any practitioner. In my view it is ignorance and confusion combined with media exposure that leads to people taking the easy option or not search for a more satisfactory lifestyle options or resolutions to health problems. If I knew at the age of 20yr what I know now I would be in much better condition rather than just falling over other therapies in desperation or by chance.

- The survey made me dig deep and my conscience kept stabbing me, but the money to cover costs of health care is running low. Princess Margaret hospital is on the brink of closure and we needed that during the earthquake. They are thinking of placing the elderly out at Burwood, but that was not a safe place during the quake. It is built on sand. What happens when there is no place left to go? Christchurch Public Hospital would never have handled the disaster if it the rupture line had been through the centre of the city. Let's keep our options open as long as we can.
- They were hard choices to make. I believe health services should be available to all and should have criteria to follow.
- I believe that healthcare is a right for all. Decisions such as these need be left to Doctors based on need. As a lay person, any decisions I would make are based on emotion and no experience.
- With health I think it is best to be dealt with on a case by case basis rather than lumping a whole lot of patients into one category as every person on this earth is different although the treatment required maybe the same the personal circumstance of each patient is different.
- I don't think you can play God in treating your patient. Whilst expert doctors may have a view and preference, you should treat them equal.
- Every case is unique and should be assessed individually (ideally).
- Interesting moral dilemmas
- The stance I took is based on all people being equal and therefore entitled to health care as a right. This is what taxes are for.
- I believe that the NZ promise was free healthcare for all; therefore, it should be available without prejudice to all. The only choices would be as far as I can see the same as for any one doctor with two patients.
- Fairness is the critical cement that holds together a civil society. I believe in the spread of a common wealth and an open government that is representative of the people and not corporations. Citizenship is comprised of rights and responsibilities and among those rights should be healthcare to all. Having lived a substantial portion of my life in the United States I can categorically say that private health care relies upon and promotes a disparity between haves and have-nots. Despite all the propaganda the U.S. is not a fair society, therefore, its discordancy will ultimately be its own undoing.
- Our health system / prioritisation is very poor. I have known a couple of people who have had serious illnesses or cancer and have had to go private as they had insurance. To find out that people with the same level of illness went through our public health system, of course from a different region, not sure how this works or is fair.
- Hopefully NZer's will halt the war on the under classes so that all can have equal access to care. Surveys like this are full of danger when assumptions are made that we have to make these choices, when the real choice was to share the nation's wealth fairly in the

first place.
- Concern about who decides which is the best form of treatment. What one profession sees as only one answer to a condition - and in cases where that answer is drastic - can have drastic side effects.
- Medics should broaden their gaze to consider 'alternative' therapies, ie massage, as a tool to better health management. Also promote education on personal responsibility for health. Prevention is better than cure.
- I have concern over unnecessary surgery. It may be advised to be the best option, when that information is misinformed.
- I have recently had cancer, used the private system for surgery, the public system for chemo, both were fantastic!
- I feel the agenda of this survey is to enable selection of patient care in the future to age of the patient. I also feel that if the hospitals weren't so top heavy financially the healthcare of each individual would benefit enormously.
- Having been through the health system a couple of times, including periods in hospitals my observation is that medical professionals have removed themselves from the human aspect of care and hide behind notions and models of customer service in a cynical fashion, so that you feel like you are an inconvenience, especially in the process of being bumped on a prioritisation list. The process could be managed better and customer service could be applied without cynicism.
- Our Health System appears to be reasonably fair to all citizens. Hospitals and health boards appear to have an abundance of administrators and I wonder if they are all needed. In the real world (don't include nurses and medical staff) I have no doubts that staff numbers would be smaller - is there a model or template that office staff levels are arrived at? Private hospitals appear to have much lower office/administrator staff levels.
- I feel if you need an operation you shouldn't have to wait more than 3 months as this puts stress on you and the person who needs the operation. Also I feel ACC are not fair on the decision they make. My husband has been denied help from them for two medical issues. We are both working people and this doesn't seem to matter to them. In both cases they were work related injuries. We both pay taxes and my husband pays ACC fees weekly but this doesn't seem to matter. I feel this is not fair. I have heard of other people who don't work who have had the help from them but they won't help us.
- Need better management of funds so more Doctors.
- Good management should bring with it efficiency and productivity.
- Data suggests doctors and nurses productivity has dropped hugely in the last few years. Why is this, and at what cost to those who need treatment. With an aging population and the number of medical professionals emigrating the present standard will be hard it maintain.
- There appear to be prescription biases in favour of giant pharmaceutical companies. Are we being held to high cost ransom by them over some common and complex medicines?
- Also the withdrawal of school dental clinics and school nurses could lead to unchecked health problems further down the line.

Own Assumptions/All else equal

- Presume that treatment for (left) will not result in (right) being denied treatment in the near future (age vs need).
- Seems better to treat someone who has their life ahead of them (patient's health vs lifestyle, didn't mention age).
- Assumed left patient receives the alternative though less effective treatment to stay alive with reasonable quality of life (treatment options vs need, didn't mention individual

benefit).
- Assume patient right will have good life after treatment.
- Assume older patient will return to productive contribution to society.
- Presume that neither patient's illness or injury is immediately life threatening.
- Doesn't say whether a full recovery and a "young" or active contributor to society.
- Based on 65+ being in good health otherwise; younger person could be long term burden on health /welfare system.
- I am assuming that neither patient's condition is not life threatening and that delaying treatment will not cause permanent negative effect.
- I don't seem to have a heart or any compassion in this survey, but in the long run the right patient will probably cost less thereby benefit society in the long term (doesn't mention cost).
- I would assume outcome for the 15-64 age bracket would generally be better than the 65+ bracket. (age vs health status)
- I think there needs to be more information to make a choice in this situation - ie how poor is poor? I used the eg of a kidney transplant.

Joint Factor Independence

- I feel as though for this question it depends on the ages and what the injury or illness is :)
- I feel as though this one also has to do with the patient's age, quality of life, what the illness is, the percentage of will the treatment help and what it is they are wanting to do :)
- This is hard because it may depend on what the treatment is and what affect any delay has on the other person.
- Difficult to assess the benefits to others without knowing further details. On the balance of probabilities the younger patient may have a large future (potential) benefit to family or society, but the older patient's current benefit to others may be more pressing. I tend to favour the future potential, but the nature of the illness or injury and whether or not it is life threatening or not would also affect my final choice.
- Depends on what the condition is and how quickly it may be treated.
- Relative ages of patients may cause me to change this decision. Also the current effect of the illness or injury on the life styles of the two patients, e.g. minor or major effects.

Making choices

- I make my choices on the basis of results. This means I may be slanting my choices towards older people when I very much want to increase the emphasis in the health system on the young.
- If the benefit of receiving the treatment to the left hand individual is medium I am prepared to give primary weight to saving a life, despite the large benefit of treatment to the other individual.
- If the benefit to the patient is large it should not depend on benefiting family or society.
- This is a hard one. As a child I would want to help my parents, but also as a mother I would want my child to healthy and have a long life (age vs health status).
- While both patients need the treatment it is difficult to weigh up which person would get the treatment as with the child there is more chance of a longer healthier life but with the elderly patient this chance is limited.
- Were rather broad questions that I don't really see the point of. Who needs the treatment and benefits the most should get the treatment. If there were more specific examples then the choices made may be different.

- Interesting exercise. I kept feeling guilty about whether or not I should be defending older people because I am older (but very healthy and physically fit) and don't see a Doctor more than 6-7 times a year (I put in once a month because as a whanau, we would average that since I have two mokos staying with me as well as my daughter - who incidentally, like my wife, doesn't get into exercise, which I have done all my life.
- I found it depended a lot on the context. I don't think age should come into it. I believe the effect of the treatment on the patient and the patient's need for it is an important factor. I also think that when treatment is needed as a result of chosen lifestyle in defiance of sound medical advice then there should be a cost to the patient or the promoters of the lifestyle.
- It was hard to make assumptions on the people. In some instances where the benefit to family / society would be medium/large - in which way? There are people with large families that don't benefit society and I feel that I would not put them ahead of those that do work and contribute to society and yet may only have small families or ahead of those in the retired age group that had worked and contributed to society all their life.
- I don't feel that a decision about who receives priority health care can be based on age or benefit to society. Needs should be the priority for any decisions made.
- I think that the main priority should be the comfort of the patient both mentally and physically and the ease of management of their ailment by themselves and others who care for them. I am not in favour of extraordinary efforts to save someone from dying only to leave them a burden to themselves and others who will tend them.
- I believe we should treat all patients equal. Sometimes, their additional life span by the treatment may not be much but it may mean a lot for their family.
- More chance of this patient being more use to family and friends and society (large benefit to society large/relatively good health > small benefit to society/die soon)
- Whilst I found this fairly easy to make my decisions it is in part because you can take away also the personalisation that is involved when working face to face with people in the health system. You also need to make a decision about what the terms fair, poor etc mean to you when deciding who has preference. Will be interested to see your results, and how this might impact decision makers within the MOH.
- I think peoples' experience with health issues eg whether they or a family member has had cancer or suffered a serious road accident, will influence the way they answer the questions. Also, the age of the respondent and whether they have children or not will influence their responses.

Difficulty of trade-off

- It was hard to answer trade off questions because every life has value and potential and it is hard to stand as judge as to who deserves or who is more valuable so who should receive care first or more of it.
- I found the health prioritisation segments quite difficult in that one cannot easily distinguish between who deserves treatment first, based on age (young versus old) and whether people should be penalised for making poor lifestyle choices (as a result of poor living conditions, opportunities etc).
- I found these questions appeared repetitious and had to think hard and long before attempting to answer then logged off and slept on it before I recommenced.

Cost

- Depends on the cost and benefits of treatment as to which patient should be treated first.
- Assuming cost benefit of treatment is favourable.
- Again I favour the option with the more significant outcome. But I would have liked to know if the treatment is an expensive one. If it was then there would be a justification for treating several people in the left box rather than one in the right hand one.
- Again there is no information about outcomes. Nor is there information about costs. If the costs were high and the benefits of medium importance I would choose the left box. If the costs were low and the benefits to an older person high I would chose the left box - a cost benefit approach.
- Cost vs Benefit doesn't seem to be involved?
- There was no choice involving costs.

Effectiveness/Efficacy

- Assuming "still beneficial" is of clinical significance and proven.
- Evidence of efficacy, NNT, NNH, cost effectiveness, patient choice, co-morbidities and treatment goals are all important in health prioritisation.
- I am quite comfortable with Health Care Service prioritization being based on analysis of controlled randomized trials.

Other/personal

- I am very pleased I don't have to make these decisions for real!
- Survey insightful.
- Thank you.
- Kia ora rawatu koe mo o mahi. Kia kaha.
- I found it a little confusing at first, but I think what you are doing is an excellent idea as we need to have a good health service for all.
- Good luck. How will we know the outcomes of your research?
- Will your thesis be available for public viewing?
- Well, thank you for the survey.
- I have worked as a caregiver to people who have come out of Porirua Hospital for about 8 years and in that time lived in a flatting situation with 8 of this group for three years and worked with people with intellectual disabilities for 12 years. I retired last Christmas and am 70 years young. I was a school teacher before that for 27 years.
- I was a radiographer before retiring in 2008.
- I have worked in the field of Disabilities and have found that when some of them enter hospital for treatment staff find them hard to understand in particular those that are hearing impaired. I have studied the NZSL so that at least they may be understood in times of need. People with disabilities are a joy to work with BUT can be a handful in the unfamiliar places. Either in homes or community based care. One in five have a disability in NZ and the deaf community have the hardest of them all.
  I worked in health care part or full time for 30 years, until eleven years ago.
- These decisions were quite difficult for me. I have had severe health issues for thirty years and have just suffered another life changing event. I have poor quality of life and yet feel very strongly that my family still need me and that I have a lot to offer.
- It would be nice to have a GP that I can afford to go to in the area that we live. I have to travel 20 to 30 mins drive to our doctors.

~ Chapter 7 ~

Do respondents' characteristics predict preferences?

7.1 Introduction

In the last chapter, the criteria weights obtained from the decision survey were presented and discussed. In this chapter the variation in the criteria weights from the random sample are analysed with respect to the demographic characteristics of the respondents to assess whether respondents' characteristics can predict preferences.[148]

The chapter begins with a discussion on modelling respondents' preferences. The rationale for using regression analysis to analyse the criteria weights is then explained, and the results are presented and discussed. The chapter concludes with a discussion relating to the regression diagnostics.

7.2 Modelling preferences

As discussed in the previous chapters, to estimate the relative importance of the six criteria established in the first part of the thesis ('need', 'individual benefit', 'age', 'societal benefit', 'lifestyle' and 'no alternative treatment') three samples of respondents (random, snowball and health services researchers) completed an online decision survey. Respondents were asked to choose between two hypothetical patients who differed on just two criteria with one patient 'described' at a higher level on one criterion and at a lower level on the other criterion compared to the other patient. As explained in Chapter 2, given the ordinal choices made by the respondents, a linear programme is implemented through 1000Minds software to obtain point 'values' or 'weights' for each respondent (Ombler & Hansen 2012). These point 'values' or 'weights' represent the relative ('part-worth') utilities of each level of each criterion.

[148] As explained in the Chapter 5, the random sample is more demographically representative of the New Zealand population than the snowball sample. As I want to assess whether the characteristics of a group of demographically representative respondents' can predict preferences, it is more appropriate to use only the random sample criteria weights.

Specifically, each respondent has a utility function (representing their preferences) which can be expressed as:

$$U_i = \overline{U_i} + U_i[f_{ia}(a) + f_{ib}(b) + \ldots + f_{ij}(f)]$$

Where U_i is the utility function for individual i, $\overline{U_i}$ is the utility from everything else, a, b,...f are the criteria and f(·) is linear and positive monotonic in the levels of each criterion.[149]

It is possible that additional criteria (apart from the six criteria mentioned above) could be included in a respondent's utility function. However, as the six criteria were established by conducting focus groups, talking to health experts and reviewing the literature, it seems reasonable to assume that the above utility function adequately represents the preferences of a representative New Zealander with respect to prioritising health treatments.

The part-worth utilities of the criteria can be compared for a respondent but they cannot be compared across respondents. For example, if the criteria weights for 'need' and 'individual benefit' are 0.4 and 0.2 respectively for Respondent A, then Respondent A considers 'need' to be twice as important as 'individual benefit'. However, relative utility of a criterion cannot be compared across respondents as utility is not a cardinal measure. For example, if the criteria weight for 'need' is 0.4 for both Respondent A and Respondent B it cannot be said that both respondents gain the same utility from 'need'. A criterion weight of 0.4 might be considered relatively high to Respondent A but it may be considered relatively low to Respondent B.

7.2.1 Analysing preferences

The criterion weights from the decision survey can be analysed in two ways. First, the MRS (or relative criteria weights) can be calculated for each respondent (or the *mean* MRS for the entire sample) and the ratios compared across respondents. The MRS is the rate at which a respondent is prepared to give up one criterion in exchange for another criterion while maintaining the same level of utility. For example, using the mean criteria weights for the random sample, displayed in Table 7.1, it can be seen that the mean MRS between 'need' and 'age' is 2 (i.e. 28.4/14.2) or conversely that the mean MRS between 'age' and 'need' is 0.5 (i.e. 14.2/28.4).

[149] Positive monotonic utility: $a_{k+1} - a_k, \geq 0 \; \forall$ a through f; where a through f are criteria and k is number of levels

Table 7.1: Random sample mean criteria weights

Criteria	Mean sample criteria weights (*N*=322)
Need	28.4%
Individual benefit	22.0%
Age	14.2%
Lifestyle	12.8%
Societal benefit	12.1%
No alternative treatment	10.5%

Second, the criteria weights can be compared across respondents bearing in mind that for each respondent the preference for one criterion is *relative* to the preference for the other criteria for that respondent (i.e. the criteria weights for each respondent add to one: $U_{respondent} = U_{need} + U_{individualbenefit} + U_{age} + U_{societalbenefit} + U_{lifestyle} + U_{noalternativetreatment}$). For example, the criterion weight for 'need' for Respondent A is 0.4 and for Respondent B it is 0.2. We can say that compared to Respondent B, Respondent A, considers 'need' to be more important relative to all of the other criteria.

Before discussing the regression techniques used in this thesis the descriptive statistics are presented and discussed.

7.3 Descriptive statistics

When respondents completed the decision survey, they were asked a series of questions relating to their demographic characteristics. This information was collected for two reasons. The first reason is to compare the demographic characteristics of the respondents with New Zealand population statistics to assess whether the samples are demographically representative. Second, the demographic characteristics of the respondents are used to assess whether respondents' characteristics can predict preferences.

As discussed in Chapter 5, in terms of demographic representativeness, the random sample more closely aligns with national statistics than the snowball sample. Therefore only the weights from the random sample will be analysed.

The demographic characteristics of the random sample are displayed in Table 7.2. Some of the demographic groups (region, age, ethnicity, employment, income, household composition and health care usage) are combined into smaller groups for analysis to avoid having groups with too few observations. For example, ethnicity information is collected in nine groups in the survey but is combined to form three groups (‘European’, ‘European/Maori’, ‘Other’) for analysis.

Table 7.2: Descriptive statistics of the independent variables (demographic characteristics)

Demographic characteristics		Number	Percentage
Gender	Male	130	40.4%
	Female	192	59.6%
Ethnicity	European	275	85.4%
	European/Maori, Maori	27	8.4%
	Other	20	6.2%
Region	Auckland	91	28.26%
	Bay of Plenty	20	6.21%
	Canterbury	48	14.91%
	Gisborne/Hawke's Bay	20	6.21%
	Manawatu-Wanganui/Taranaki	19	5.90%
	Marlb./Nelson/Tasman/West Coast	17	5.28%
	Northland	12	3.73%
	Otago/Southland	27	8.39%
	Waikato	29	9.01%
	Wellington	39	12.11%
Age	18-34 yrs	54	16.8%
	35-54 yrs	116	36.0%
	55 and over	152	47.2%
Income	$0-$30,000	64	19.9%
	$30,001-$70,000	101	31.4%
	Over $70,000	108	33.5%
	Not given	49	15.2%
Qualifications	No qualifications	19	2.9%
	Secondary school	107	33.2%
	Other post secondary school quals	83	25.8%
	University degree or equivalent	113	35.1%
Household composition	Not living with children	193	60.1%
	Living with children	128	39.9%
Employment	Working	201	62.4%
	Not working	49	15.2%
	Retired	63	19.6%
Worker type	Health related worker	33	10.2%
	Non-health related worker	289	89.8%
Experience of serious illness	Yes	211	65.5%
	No	111	34.5%
Health insurance	Yes	143	44.4%
	No	179	55.6%
Health use	Seldom	58	18.0%
	Occasionally	163	50.6%
	Frequently	99	30.7%

As mentioned above, the demographic information can be used to estimate whether respondents' characteristics can predict preferences. Many of the demographic characteristics listed in Table 7.2 are associated with the criteria in some way. For example, Maori have higher health needs than non-Maori (Ministry of Health 2011). Therefore, can we expect Maori, on average, to place a relatively higher value on 'need' compared to non-Maori? People who are better-educated and earn higher incomes live longer than people who are less-educated and are on low incomes (Deaton 2003). Socioeconomic status is also found to be correlated with 'risky' behaviours such as smoking, binge drinking, obesity and lack of exercise (Deaton 2003). Will respondents who are relatively more educated, in full-time employment or who earn more than other respondents place more value on 'need' or 'societal benefit' and less value on 'age' or 'lifestyle' compared to less-educated, unemployed or low-income respondents? What effect will the age of a respondent have on the criterion 'age'? Will older respondents favour the young or favour the old? Does having health insurance or the frequency of seeking health services affect the relative value of any of the criteria?

All the demographic characteristics listed in Table 7.2 are included in the regression analyses (explained in the next section).

7.4 Regression analysis[150]

Regression analysis is used to estimate whether the demographic characteristics of the respondents can explain the heterogeneity in the criteria weights. For example, what effect would a change in one of the demographic characteristics (predictor variable) have on the mean weights for one (or more) of the criteria (outcome variable)? As discussed in Section 7.2, this can be done in two ways: MRS between every pair of criteria for each respondent are regressed against the demographic characteristics of the respondents *or* the criteria weights can be compared across respondents, taking into account that individual criteria weights are *relative* to the other criteria weights for each respondent. Both ways are used in this thesis.

7.5 Regressions using MRS

The model used for the MRS regressions is Ordinary Least Squares (OLS). OLS is a simple linear regression model that is used to estimate slope parameters by minimising the squared

[150] The statistical software package 'Stata 11" was used for the regression analyses.

vertical differences between observed values and predicted values. The MRS between each pair of criteria are calculated for each respondent and then regressed against the demographic characteristics of the respondents. As there are six criteria, there are 15 regressions (i.e. n(n-1)/2 where n=number of criteria). Because of the number of regressions the results have been split into three tables (Tables 7.3, 7.4 and 7.5).

Due to the low response rate and potential non-coverage bias (discussed in Chapter 5), a conservative approach is taken in interpreting the significant results. Only the demographic characteristics that have coefficients which are statistically significant at the 5% or 1% level will be discussed. In addition, though some of the 'region' coefficients are statistically significant, they are being compared with one reference category only. As interpretation of the statistically significant 'region' coefficients is of little value to the analysis, they will not be discussed.

Apart from 'age group' all the other demographic groups are dummy variables.[151] This means that the coefficients can be interpreted as the 'average' change in the ratio being considered, given a change in the level of the demographic characteristic (holding all other variables constant). For example, as can be seen Table 7.3 below, in the second column titled 'only treatment/individual benefit', on average, respondents who have an annual household income greater than $70,000 have a lower MRS (on average 32% lower) between 'only available treatment' and 'individual benefit' compared to respondents who have an annual household income between $30,001 and $70,000 (statistically significant at the 5% level). In this example, a lower MRS means that the higher income group is willing to give up 'one unit' of 'only available treatment' for a smaller amount of 'individual benefit' compared to respondents in the middle-income group or put simply, the higher income group places less importance on 'only available treatment' relative to 'individual benefit' compared to the 'middle- income' group.

[151] The demographic groups are coded into dummy variables (or multiple dummy variables) for regression analysis. A dummy variable represents a non-numeric variable and takes a value of 0 or 1; for example, 0 for 'females' and 1 for 'males'. The demographic groups that have more than two categories are recoded into multiple dummy variables; for example, 1 for 'seldom health use', 2 for 'occasional health use' and 3 for 'frequent health use'.

Table 7.3: Comparison of MRS across respondents

Demographic Characteristics (N=322)	Societal benefit/ individual benefit	Only treatment/ individual benefit	Need/ individual benefit	Need/ lifestyle	Only treatment/ need
Male	0.00621	−0.0246	0.0835	**1.507****	**−0.0984*****
(ref: female)	(0.0588)	(0.0690)	(0.135)	(0.701)	(0.0374)
European/Maori, Maori	0.00924	−0.0347	0.190	0.431	−0.0716
(ref: European)	(0.105)	(0.117)	(0.241)	(1.239)	(0.0529)
Other ethnicity	−0.0666	0.0410	0.152	−0.410	−0.0216
(ref: European)	(0.145)	(0.150)	(0.350)	(0.940)	(0.0646)
Age (continuous)	−0.0214	0.0130	−0.00570	−0.319	0.0205
Mean age 45-54 years	(0.0306)	(0.0329)	(0.0625)	(0.280)	(0.0164)
Income $30,001-$70,000	−0.0870	−0.247*	−0.347	−0.609	−0.0848
(ref: $0-$30,000)	(0.0885)	(0.142)	(0.269)	(1.040)	(0.0711)
Income over $70,000	−0.114	**−0.320****	**−0.704*****	−0.890	−0.0692
(ref: $0-$30,000)	(0.0874)	(0.145)	(0.245)	(1.185)	(0.0750)
Income over $70,000	−0.0270	−0.0726	**−0.358****	−0.280	0.0156
(ref: $30,001−$70,000)	(0.0867)	(0.0764)	(0.156)	(0.852)	(0.445)
Not living with children	0.0537	0.202*	**0.451****	0.000954	0.00360
(ref: living with children)	(0.0795)	(0.106)	(0.220)	(0.698)	(0.0378)
Occasional health use	−0.0489	**−0.169****	**−0.360****	−0.650	−0.0129
(ref: seldom health use)	(0.0658)	(0.0765)	(0.152)	(0.707)	(0.0443)
Frequent health use	0.104	0.0281	0.201	−0.0139	−0.0694
(ref: seldom health use)	(0.117)	(0.143)	(0.289)	(1.032)	(0.0500)
Frequent health use	0.153	0.198*	**0.562****	0.636	−0.0565
(ref: occasional health use)	(0.107)	(0.119)	(0.253)	(0.905)	(0.0426)
No serious illness	0.0601	0.0195	−0.0483	0.296	0.0227
(ref: serious illness)	(0.0688)	(0.0693)	(0.143)	(0.660)	(0.0368)
No health insurance	0.0307	−0.0160	−0.0318	−0.929	0.0607
(ref: health insurance)	(0.0639)	(0.0645)	(0.134)	(0.619)	(0.0411)
Sec school qualifications	0.0380	−0.205	−0.337	−0.154	0.0270
(ref: no qualifications)	(0.113)	(0.311)	(0.600)	(1.543)	(0.0647)
Other post sec school quals	0.0319	−0.207	−0.311	−0.417	0.0157
(ref: no qualifications)	(0.107)	(0.294)	(0.571)	(1.483)	(0.0668)
Degree or equivalent	0.0445	−0.157	−0.399	−0.0154	0.0612
(ref: no qualifications)	(0.108)	(0.284)	(0.552)	(1.511)	(0.0694)
Not working	**−0.153****	−0.0366	−0.243	0.538	−0.00168
(ref: working)	(0.0777)	(0.0961)	(0.185)	(0.830)	(0.0539)
Retired	−0.0904	−0.0129	−0.187	0.624	0.00736
(ref: working)	(0.0766)	(0.106)	(0.195)	(1.308)	(0.0577)
Retired	0.0626	0.0236	0.0564	0.0856	0.00905
(ref: not working)	(0.102)	(0.126)	(0.216)	(1.356)	(0.0832)
Health worker	−0.0801	0.0106	0.186	0.800	−0.0501
(ref: non-health worker)	(0.136)	(0.0873)	(0.234)	(0.769)	(0.0646)
Region2	0.0401	−0.138	−0.470***	−0.197	0.107
	(0.135)	(0.0904)	(0.169)	(1.543)	(0.0830)
Region3	−0.0280	0.0676	0.288	−0.263	−0.0149
	(0.0917)	(0.0921)	(0.211)	(0.713)	(0.0507)
Region4	0.0281	0.266	0.501	−1.522**	−0.0115
	(0.120)	(0.246)	(0.470)	(0.728)	(0.0719)
Region5	0.0679	0.196	0.735*	1.966	−0.0657
	(0.177)	(0.174)	(0.431)	(2.324)	(0.0534)
Region 6	0.148	0.457**	0.380	2.371	0.247***
	(0.177)	(0.183)	(0.342)	(2.281)	(0.0947)
Region 7	−0.00339	0.183	0.386	−0.330	0.0575
	(0.113)	(0.152)	(0.339)	(1.051)	(0.0944)
Region 8	−0.00208	0.0117	0.450*	0.671	0.0282
	(0.148)	(0.0830)	(0.251)	(0.983)	(0.114)
Region 9	0.0724	0.199	0.530**	1.897	−0.0347
	(0.157)	(0.135)	(0.257)	(1.572)	(0.0571)
Region 10	−0.0171	0.0545	0.107	1.500	0.0251
	(0.0938)	(0.114)	(0.230)	(1.107)	(0.0545)
Constant	0.708**	0.902***	2.030***	4.172*	0.434***
	(0.285)	(0.342)	(0.693)	(2.283)	(0.134)
R−squared	0.047	0.134	0.165	0.085	0.113

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 7.4: Comparison of MRS across respondents

Demographic Characteristics (N=322)	Only treatment/age	Age/ societal benefit	Age/ need	Age/ individual benefit	Age/ lifestyle
Male	−0.0973	−0.248	−0.0180	−0.0773	0.358
(ref: female)	(0.138)	(0.314)	(0.0688)	(0.117)	(0.396)
European/Maori, Maori	0.305	−0.0845	**−0.229*** **	−0.175	−0.174
(ref: European)	(0.256)	(0.724)	(0.0787)	(0.176)	(0.854)
Other ethnicity	0.249	1.515	−0.120	−0.0727	−0.285
(ref: European)	(0.242)	(1.169)	(0.0898)	(0.235)	(0.476)
Age (continuous)	0.121*	**−0.407****	−0.0333	**−0.0836****	−0.193
Mean age 45-54 years	(0.0680)	(0.160)	(0.0266)	(0.0411)	(0.133)
Income $30,001-$70,000	0.0640	−0.0720	0.0146	0.0211	−0.219
(ref: $0-$30,000)	(0.236)	(0.448)	(0.0801)	(0.186)	(0.662)
Income over $70,000	−0.156	0.340	0.102	−0.167	−0.186
(ref: $0-$30,000)	(0.220)	(0.626)	(0.0915)	(0.125)	(0.685)
Income over $70,000	−0.220	0.412	0.0875	−0.189	0.0331
(ref: $30,001-$70,000)	(0.185)	(0.453)	(0.0853)	(0.154)	(0.484)
Not living with children	0.0712	−0.177	−0.0202	0.197	−0.344
(ref: living with children)	(0.137)	(0.368)	(0.0729)	(0.125)	(0.365)
Occasional health use	0.159	−0.754	−0.0985	**−0.330****	−0.711
(ref: seldom health use)	(0.140)	(0.458)	(0.103)	(0.148)	(0.445)
Frequent health use	−0.122	−0.491	−0.136	−0.123	0.0507
(ref: seldom health use)	(0.139)	(0.566)	(0.114)	(0.168)	(0.888)
Frequent health use	**−0.281****	0.263	−0.0374	0.207*	0.762
(ref: occasional health use)	(0.139)	(0.410)	(0.0653)	(0.107)	(0.647)
No serious illness	**−0.307****	0.0200	0.0751	0.123	**0.811****
(ref: serious illness)	(0.155)	(0.400)	(0.0880)	(0.0992)	(0.374)
No health insurance	0.0277	−0.552*	0.0746	0.00720	−0.230
(ref: health insurance)	(0.136)	(0.328)	(0.0639)	(0.116)	(0.352)
Sec school qualifications	−0.276	0.792	0.117	0.157	−0.997
(ref: no qualifications)	(0.376)	(0.728)	(0.137)	(0.215)	(1.673)
Other post sec school quals	−0.258	0.196	0.0299	−0.0642	−1.493
(ref: no qualifications)	(0.370)	(0.675)	(0.136)	(0.193)	(1.681)
Degree or equivalent	−0.0884	0.0786	−0.0291	−0.0624	−1.467
(ref: no qualifications)	(0.368)	(0.665)	(0.132)	(0.193)	(1.739)
Not working	0.420*	−0.457	−0.125*	**−0.335****	−0.190
(ref: working)	(0.238)	(0.517)	(0.0687)	(0.135)	(0.431)
Retired	0.242	0.381	0.0449	0.0106	0.638
(ref: working)	(0.234)	(0.381)	(0.101)	(0.118)	(0.687)
Retired	−0.178	0.838	0.170	**0.345****	0.828
(ref: not working)	0.342)	(0.660)	(0.117)	(0.157)	(0.726)
Health worker	0.0589	0.334	−0.103	0.0131	0.307
(ref: non-health worker)	(0.150)	(0.458)	(0.0887)	(0.145)	(0.356)
Region2	0.0469	−0.552	0.125	−0.189	0.125
	(0.349)	(0.352)	(0.141)	(0.164)	(0.852)
Region3	−0.0357	−0.102	−0.106	−0.0568	0.105
	(0.219)	(0.397)	(0.105)	(0.211)	(0.673)
Region4	0.179	0.0244	−0.126	0.0709	−0.742
	(0.290)	(0.602)	(0.128)	(0.279)	(0.475)
Region5	−0.0624	0.285	−0.167*	0.0297	0.418
	(0.237)	(0.623)	(0.0935)	(0.249)	(0.749)
Region 6	1.066***	−0.919**	−0.236	−0.363	−0.472
	(0.359)	(0.454)	(0.154)	(0.270)	(0.499)
Region 7	−0.344	2.760	0.227	0.189	0.127
	(0.262)	(2.061)	(0.196)	(0.188)	(0.662)
Region 8	−0.132	0.812	−0.0789	0.0676	0.513
	(0.253)	(0.787)	(0.108)	(0.240)	(0.553)
Region 9	−0.309*	0.149	−0.0571	0.165	0.747
	(0.176)	(0.439)	(0.0955)	(0.232)	(0.569)
Region 10	−0.280	0.612	0.0750	0.114	2.101**
	(0.196)	(0.584)	(0.113)	(0.207)	(0.961)
Constant	1.147*	4.449***	0.715***	1.312***	2.669
	(0.691)	(1.641)	(0.234)	(0.500)	(2.055)
R−squared	0.125	0.138	0.088	0.099	0.115

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 7.5: Comparison of MRS across respondents

Demographic Characteristics (N=322)	Only treatment /societal benefit	Only treatment /lifestyle	Societal benefit/ need	Societal benefit/ lifestyle	Individual benefit/ lifestyle
Male	−0.375	0.166	−0.0196	0.558*	0.824*
(ref: female)	(0.242)	(0.334)	(0.0529)	(0.320)	(0.492)
European/Maori, Maori	0.151	−0.194	−0.101	0.168	−0.226
(ref: European)	(0.466)	(0.455)	(0.0778)	(0.793)	(1.059)
Other ethnicity	0.900	−0.00287	−0.159*	−0.484	−1.221*
(ref: European)	(0.718)	(0.490)	(0.0828)	(0.499)	(0.639)
Age (continuous)	−0.234	−0.0378	0.0138	−0.167	−0.0755
Mean age 45-54 years	(0.142)	(0.102)	(0.0213)	(0.157)	(0.181)
Income $30,001-$70,000	−0.164	−0.751	−0.0601	−0.329	−0.657
(ref: $0-$30,000)	(0.431)	(0.630)	(0.0649)	(0.435)	(0.798)
Income over $70,000	−0.183	−0.611	0.106	−0.188	−0.533
(ref: $0-$30,000)	(0.541)	(0.727)	(0.0763)	(0.452)	(0.820)
Income over $70,000	−0.0198	0.140	**0.166****	0.140	0.124
(ref: $30,0001-$70,000)	(0.298)	(0.369)	(0.0687)	(0.386)	(0.628)
Not living with children	−0.142	0.0380	−0.0211	−0.0794	−0.150
(ref: living with children)	(0.330)	(0.236)	(0.0584)	(0.368)	(0.524)
Occasional health use	−0.560*	−0.606*	−0.0430	−0.123	−0.0623
(ref: seldom health use)	(0.302)	(0.341)	(0.0817)	(0.354)	(0.552)
Frequent health use	−0.0936	−0.225	−0.102	−0.0697	−0.216
(ref: seldom health use)	(0.467)	(0.480)	(0.0931)	(0.463)	(0.818)
Frequent health use	0.466	0.381	−0.0588	0.0536	−0.154
(ref: occasional health use)	(0.387)	(0.366)	(0.0536)	(0.418)	(0.658)
No serious illness	−0.254	0.218	0.0278	0.542*	0.334
(ref: serious illness)	(0.304)	(0.279)	(0.0653)	(0.276)	(0.493)
No health insurance	−0.361	−0.0335	0.0569	−0.0635	−0.0136
(ref: health insurance)	(0.235)	(0.334)	(0.0543)	(0.316)	(0.475)
Sec school qualifications	0.196	−0.826	0.0690	0.0967	−0.549
(ref: no qualifications)	(0.511)	(0.923)	(0.0795)	(0.576)	(1.182)
Other post sec school quals	0.0111	−0.795	0.0790	0.0958	−0.655
(ref: no qualifications)	(0.478)	(0.925)	(0.0803)	(0.579)	(1.161)
Degree or equivalent	0.251	−0.291	0.0497	0.428	−0.126
(ref: no qualifications)	(0.563)	(0.988)	(0.0796)	(0.666)	(1.251)
Not working	0.278	0.318	−0.0518	−0.271	0.160
(ref: working)	(0.351)	(0.372)	(0.0648)	(0.340)	(0.576)
Retired	0.584	0.809	−0.0165	0.267	0.288
(ref: working)	(0.372)	(0.646)	(0.0804)	(0.369)	(0.689)
Retired	0.306	0.490	0.0353	0.538	0.129
(ref: not working)	(0.485)	(0.711)	(0.0972)	(0.481)	(0.857)
Health worker	0.222	0.340	−0.179*	−0.112	0.259
(ref: non-health worker)	(0.384)	(0.369)	(0.103)	(0.426)	(0.616)
Region2	−0.419	−0.0441	0.336**	0.306	0.109
	(0.347)	(0.537)	(0.169)	(0.580)	(1.058)
Region3	−0.150	−0.207	−0.100	−0.265	−0.762
	(0.334)	(0.261)	(0.0809)	(0.457)	(0.712)
Region4	−0.00433	−0.487	−0.0865	−0.577	−1.730***
	(0.344)	(0.345)	(0.0846)	(0.351)	(0.649)
Region5	0.690	0.837	−0.0603	0.196	−0.308
	(0.677)	(1.029)	(0.0974)	(0.558)	(1.170)
Region 6	0.852	1.250*	0.0219	0.756	−0.336
	(0.628)	(0.682)	(0.138)	(1.129)	(0.985)
Region 7	0.771	−0.0620	−0.0353	−0.310	−1.118*
	(1.022)	(0.500)	(0.164)	(0.424)	(0.677)
Region 8	−0.0150	0.680	−0.112	0.325	0.556
	(0.465)	(0.940)	(0.0928)	(0.660)	(1.164)
Region 9	−0.0151	0.709	−0.122	−0.0336	−0.342
	(0.374)	(0.695)	(0.0810)	(0.419)	(0.677)
Region 10	0.352	0.383	−0.0375	0.958	1.281
	(0.615)	(0.476)	(0.0773)	(0.745)	(1.010)
Constant	3.795***	2.053*	0.551**	1.004	2.861
	(1.444)	(1.207)	(0.222)	(1.126)	(1.805)
R−squared	0.093	0.095	0.121	0.064	0.065

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The statistically significant results (at the 5% and 1% levels) are discussed below.

7.5.1 Gender

On average, male respondents have a much higher MRS between ‘need’ and ‘lifestyle’ (+150.7%, $p < 0.05$) compared to females, and a lower MRS between ‘only available treatment’ and ‘need’(−9.8%, $p < 0.001$). This means that male respondents tend to place more importance on ‘need’ relative to ‘lifestyle’ and ‘only available treatment’ compared to female respondents. It is possible that females are more aware of (or at least concerned about) the impact of poor lifestyle choices on health and the availability of treatments, compared to males.

7.5.2 Ethnicity

Respondents who identify as Maori have a lower MRS between ‘age’ and ‘individual benefit’ (−22.9%, $p < 0.001$) compared to respondents who do not identify as Maori. According to a discussion in the Maori health provider focus group, Maori consider the wellbeing of children to be very important but they also consider older people to be valuable in terms of looking after the whanau. This is perhaps why Maori place more importance on ‘individual benefit’ and less importance on ‘age’ (i.e. younger members of society) compared to non-Maori.

7.5.3 Age of respondents

Respondents over 54 years, on average, have a lower MRS between ‘age’ and ‘societal benefit’ (−40.7%, $p < 0.05$) and a lower MRS between ‘age’ and ‘individual benefit’ (−8.36%, $p < 0.05$) compared to respondents younger than 45. This means that on average, younger respondents show a greater preference for treatments that benefit young people (relative to treatments that benefit society or an individual) whereas older respondents show a greater preference for treatments that benefit individuals or society (relative to treatments that benefit the young).

7.5.4 Household income

On average, respondents with a household income over $70,000 have a lower MRS between ‘only available treatment’ and ‘individual benefit’ (−32%, $p < 0.05$), and a lower MRS

between ‘need’ and ‘individual benefit’ (–70.4%, $p < 0.001$) compared to respondents with a household income less than $30,000. They also have a lower MRS between ‘need’ and ‘individual benefit’ (–35.8%, $p < 0.05$) and a higher MRS between ‘societal benefit’ and ‘need’ (+16.6%, $p < 0.05$) compared to respondents with a household income between $30,001 and $70,000.

High-income respondents, on average, have a higher relative weight on individual benefit compared to middle and low-income respondents (this is discussed further in Section 7.7.3). This means that compared to middle-income respondents, high-income respondents place more importance on ‘individual benefit’ relative to ‘need’ and ‘societal benefit’; and, compared to low-income respondents, they place more importance on ‘individual benefit’ relative to ‘need’ and ‘only available treatment’.

7.5.5 Living with children

Respondents who do *not* live with children, on average, have a lower MRS between ‘need’ and ‘individual benefit’ (–45.1%, $p < 0.05$) compared to respondents who live with children. This indicates that respondents who live with children place relatively more importance on ‘need’ and less on ‘individual benefit’ compared to respondents who do not live with children. This result may not be surprising to parents who often put the needs of others (e.g. their children) before themselves.

7.5.6 Health usage

On average, respondents who use health services occasionally, have a lower MRS between ‘need’ and ‘individual benefit’ (–36%, $p < 0.05$), between ‘age’ and ‘individual benefit (–33%, $p < 0.05$) and between ‘only available treatment’ and ‘individual benefit’ (–16.9%, $p < 0.05$) compared to respondents who seldom use health services. This means that respondents who seldom use health services place more importance on ‘individual benefit’ relative to ‘need’, ‘age’ and ‘only available treatment’ compared to respondents who use health services occasionally.

Respondents who use health services frequently, on average, have a lower MRS between ‘no alternative treatment’ and ‘age’ (–28.1%, $p < 0.05$) and a higher MRS between ‘need’ and

'individual benefit' (+56.2%, $p < 0.05$) compared to respondents who use health services occasionally.

7.5.7 Experience of a serious illness

Respondents who have *not* experienced a serious illness, on average, have a lower MRS between 'only available treatment' and 'age' (–30.7%, $p < 0.05$) compared to respondents who *have* experienced a serious illness. This means that respondents who have experienced a serious illness place more importance on 'only available treatment' relative to 'age' compared to respondents who have not experienced a serious illness. Possibly treatment options were limited for some of the respondents who have experienced a serious illness and that is why on average, they place relatively more importance on 'only available treatment' compared to 'age'.

7.5.8 Work status

On average, respondents who are not working have a lower MRS between 'age' and 'individual benefit' (–33.5%, $p < 0.05$) and between 'societal benefit' and 'individual benefit' (–15.3%, $p < 0.05$) compared to respondents who work. Respondents who are employed therefore place more importance on 'individual benefit' relative to 'age' and 'societal benefit' compared to respondents who do not work. It may be that employed respondents are aware that good health is conductive to productivity and therefore they place relatively more importance on 'individual benefit'. In contrast, respondents who are retired, on average, place more importance on 'age' relative to 'individual benefit' compared to respondents who are not working.

7.5.9 Discussion of MRS results

Many of the results in Section 7.5 are plausible. For example, respondents who are employed place more importance on 'individual benefit' relative to 'age' or 'societal benefit' compared to respondents who do not work. Respondents who have an annual household income over $70,000 place more importance on 'individual benefit' relative to 'need' compared to respondents who have an annual household income of between $30,000-$70,000. Similarly, respondents who do *not* live with children place more importance on 'individual benefit' relative to 'need' compared to respondents who live with children, As might be expected,

respondents who have experienced a serious illness place more importance on 'only available treatment' relative to 'age' compared to respondents who have not experienced a serious illness.

The statistically significant results discussed in the previous section reveal that certain demographic characteristics influence respondents' preferences for *one criterion over another*. In the next section the mean weights for the *six* criteria are regressed against the demographic characteristics of the respondents.

7.6 Regressions using criteria weights

To explore the relationship between the demographic characteristics of the respondents and the criteria weights, OLS, Seemingly Unrelated Regressions (SUR) and the fractional multinomial logit (FML) model are used.

As mentioned in Section 7.5, OLS is a simple linear regression model that is used to estimate slope parameters by minimising the squared vertical differences between observed values and predicted values. SUR regressions are used when there are several linear equations within a single model, causing the error terms to be related. The main difference between SUR and OLS is the size of the standard error. As the standard error is smaller using SUR regression, SUR regressions produce several more significant coefficients compared with OLS.

The FML model is similar to the multinomial logit (MNL) model. The principal difference between the models is that the dependent variable in the MNL model consists of multiple categories whereas in the FML model the dependent variable consists of proportions. A requirement of the FML model is that the dependent variables (criteria weights) range between zero and one and for each observation the variables (set of criteria weights) add to one (Buis 2010).

All models (OLS, SUR and FML) produce similar results. For instance, though the level of statistical significance differs for some of the criteria, the same criteria/demographic groups exhibit statistical significance (except for 'serious illness') using OLS and FML regressions. As there are minimal differences in results between the models, and given that the FML model seems the most appropriate model in these circumstances (i.e. the criteria weights are

proportions), only the results from the FML model will be discussed. The results of the OLS and SUR regressions are in Appendices 7.1 and 7.2.

7.7 The fractional multinomial logit (FML) model

The data to be used in the FML model are as follows:

Dependent variable: criteria weights for 'need', 'individual benefit', 'societal benefit', 'age', 'no alternative treatment' and 'lifestyle' for each respondent

Regressors: demographic characteristics – gender, ethnicity, age, income, household composition, health usage, serious illness, insurance, qualifications, employment, worker type and region

The fractional multinomial logit (FML) model is an extension of the fractional logit model developed by Papke & Wooldridge (1996). The authors used the model in their study on voluntary individual contributions to retirement schemes. The dependent variable in their model was the fraction of allowable contributions each individual chooses to contribute to a retirement account. The authors found that the model provided relatively efficient estimates of the univariate conditional mean. The model has since been extended (to the FML model) and applied in several areas including commodity flows (Sivakumar & Bhat 2002), transportation time (Ye & Pendyala 2005), household time allocation (Mullahy & Robert 2010) and expenditure shares (Koch 2010).

The FML model is described as follows. Each respondent has a set of criteria weights (between the values of zero and one) that add to one:

Given $\mathrm{E}[y_{Ji} \mid x_i] \in (0 < y_{ji} < 1)$ for all *i*;

$$\sum_{j=1}^{J} \mathrm{E}[y_{Ji} \mid x_i] = 1 \text{ for all } i;$$

(where *J* = the six criteria, *x* = the demographic characteristics, and *i* = each respondent)

A functional form that encompasses these considerations is the MNL functional form (Mullahy 2010). The predicted values ($\hat{y}$) of the criteria weights depend on the demographic characteristics (x) of each respondent (i) (Buis 2010):

$$\hat{y}_{1i} = \frac{1}{\sum_{k=1}^{6} \exp(x_i \beta_k)}$$

$$\hat{y}_{2i} = \frac{\exp(x_i \beta_{k1})}{\sum_{k=1}^{6} \exp(x_i \beta_k)}$$

....

$$\hat{y}_{6i} = \frac{\exp(x_i \beta_6)}{\sum_{k=1}^{6} \exp(x_i \beta_k)}$$

where 1,2...6 are the six criteria

The MNL function is normalised by setting one of the parameters equal to zero. This means that the six criteria are regressed separately on the demographic characteristics of the respondents using one of the criteria as the 'reference' or 'omitted' category against which the other criteria can be compared (Greene 2000). The coefficients of the remaining criteria are interpreted relative to the 'base criterion'.[152]

$$\hat{y}_{6i} = \frac{\exp(x_i \beta_6)}{1 + \sum_{k=1}^{6} \exp(x_i \beta_k)}$$

where 1 ... 6 are the six criteria, $\beta_1 = 0$

Maximum likelihood estimation[153] is used to estimate the effects on the criteria weights given the demographic characteristics of the respondents and the parameters $\beta_2 \ldots \beta_6$.

The same log likelihood function (which is a function of the predicted values) is used for both the MNL and FML models:[154]

$$\ln(L_i) = y_{1i} \ln(\hat{y}_{1i}) + y_{2i} \ln(\hat{y}_{2i}) + \ldots + y_{6i} \ln(\hat{y}_{6i})$$

[152] Separate FML regressions were carried out using each criterion as the base category. The choice of base criterion made no difference to the results.

[153] Maximum likelihood estimation is a method used to estimate the parameters of a model. The method selects parameters that are the most likely to have produced the observed distribution of data (Field 2009).

[154] The predicted values in the MNL model are probabilities whereas in the FML model the predicted values are proportions.

However, with the MNL model the log likelihood function requires that each category of the dependent variable takes the value of zero or one (Koch 2010).[155] Therefore $\hat{y}$ is not included in the MNL log likelihood function unless y=1. In contrast, each category of the dependent variable in the FML model takes a value *between* zero and one (i.e. the values are continuous rather than discrete). Therefore all $\hat{y}$ are included in the FML log likelihood function. The log likelihood function is maxmised to obtain the β parameters.

Using a base criterion (i.e. $\beta_1 = 0$) to interpret the coefficients of the other criteria is not straightforward. As the primary reason for using the FML model is to estimate the effect of a change in one of the demographic characteristics on a conditional mean criterion weight, the average partial effects (APEs) are more relevant (Koch 2010, Mullahy & Robert 2010).

In Tables 7.6 to 7.8 the APEs are presented for the entire random sample, for female respondents and for male respondents, respectively. The APE measures the average change in the expected conditional mean of a criterion when there is an incremental change in a demographic group (holding all other variables constant).

How the APE is calculated depends on whether the values of a demographic group are continuous or discrete. For example, the demographic variable, 'age', is continuous. The APE is therefore computed at the mean (45-54 years).[156] As can be seen in Table 7.6 (row seven, column three) on average, respondents aged over 54 years place 1.1 percentage points[157] *less* on the criterion 'age' (relative to the other criteria) compared to respondents younger than 45 years.

As explained in Section 7.5, apart from 'age' all the other demographic groups are dummy (or discrete) variables. This means that the APE measures the average change of a conditional mean criterion weight given a change in the *level* of a demographic characteristic. For example, on average, males place 1.1 percentage points *less* on the criterion 'only available treatment' (relative to the other criteria) compared to females (Table 7.6, row one, column one).

---

[155] The dependent variable in the MNL model can be thought of as a set of dummy variables $y_0 \ldots y_J$ that take the value of either zero or one (Buis 2010).

[156] FML regressions were also run with 'age' as a dummy variable (in three groups). The differences in the results with 'age' as a continuous variable and 'age' as a dummy variable are minimal.

[157] Because the criterion weights are percentages (each set of criteria weights adds to one), 'percentage points' can be interpreted as a 'percentage'.

For each regression the model's 'goodness-of-fit' (i.e., how well the model describes the data) is assessed by using the Wald chi-square test. The Wald chi-square statistic is statistically significant at the 1% level for all of the criteria indicating that at least one of the coefficients in the model has an impact on the criteria.

Table 7.6: Average partial effects using a FML model

Demographic characteristics	Only available treatment	Need	Age	Societal benefit	Individual benefit	Lifestyle
Male	−0.011	0.016	0.001	0.004	0.002	−0.012
(ref: female)	(0.013)	(0.014)	(0.010)	(0.009)	(0.010)	(0.016)
European/Maori, Maori	−0.002	**0.044****	−0.025*	−0.003	−0.005	−0.010
(ref: European)	(0.014)	(0.019)	(0.013)	(0.013)	(0.015)	(0.014)
Other ethnicity	0.010	0.032	−0.002	−0.023	−0.013	−0.004
(ref: European)	(0.013)	(0.021)	(0.018)	(0.016)	(0.019)	(0.017)
Age (continuous)	0.003	−0.001	**−0.011*****	0.003	**0.008****	0.001
Mean age 45-54 years	(0.003)	(0.004)	(0.004)	(0.035)	(0.004)	(0.003)
Income $30,001-$70,000	−0.014	−0.003	0.005	−0.005	0.018	−0.001
(ref: $0-$30,000)	(0.010)	(0.015)	(0.013)	(0.011)	(0.015)	(0.014)
Income over $70,000	−0.021*	**−0.031****	0.003	0.009	**0.034****	0.007
(ref: $0-$30,000)	(0.011)	(0.015)	(0.013)	(0.012)	(0.015)	(0.015)
Income over $70,000	−0.008	**−0.028****	−0.002	0.014	0.015	0.008
(ref: $30,001-$70,000)	(0.009)	(0.013)	(0.010)	(0.011)	(0.012)	(0.012)
Not living with children	0.007	0.008	−0.002	−0.001	−0.016	0.005
(ref: living with children)	(0.008)	(0.011)	(0.009)	(0.009)	(0.011)	(0.010)
Occasional health use	−0.008	−0.016	**−0.020****	−0.001	**0.029****	0.016
(ref: seldom health use)	(0.008)	(0.013)	(0.010)	(0.009)	(0.012)	(0.013)
Frequent health use	−0.007	0.007	−0.010	0.005	0.010	−0.006
(ref: seldom health use)	(0.010)	(0.015)	(0.012)	(0.012)	(0.014)	(0.014)
Frequent health use	0.001	0.023*	0.010	0.006	−0.019	−0.021*
(ref: occasional health use)	(0.009)	(0.013)	(0.011)	(0.010)	(0.012)	(0.012)
No serious illness	−0.001	−0.025	0.019	0.007	−0.009	0.009
(ref: serious illness)	(0.008)	(0.017)	(0.023)	(0.012)	(0.014)	(0.016)
No health insurance	0.004	−0.009	0.004	0.003	−0.006	0.005
(ref: health insurance)	(0.009)	(0.012)	(0.009)	(0.008)	(0.010)	(0.010)
Sec school qualifications	−0.005	0.001	0.020	0.008	−0.017	−0.006
(ref: no qualifications)	(0.017)	(0.019)	(0.022)	(0.016)	(0.019)	(0.022)
Other post sec school quals	−0.005	0.007	0.003	0.013	−0.007	−0.012
(ref: no qualifications)	(0.017)	(0.020)	(0.022)	(0.016)	(0.019)	(0.022)
Degree or equivalent	0.007	0.001	0.000	0.012	−0.004	−0.017
(ref: no qualifications)	(0.017)	(0.020)	(0.021)	(0.016)	(0.019)	(0.022)
Not working	0.013	0.010	−0.020*	−0.004	0.023*	−0.021
(ref: working)	(0.011)	(0.014)	(0.011)	(0.010)	(0.013)	(0.013)
Retired	0.006	−0.017	0.010	−0.003	0.004	0.001
(ref: working)	(0.013)	(0.017)	(0.013)	(0.011)	(0.013)	(0.014)
Retired	−0.007	−0.028	0.031*	0.000	−0.020	0.024
(ref: not working)	(0.014)	(0.019)	(0.019)	(0.014)	(0.016)	(0.020)
Health worker	−0.002	**0.036****	−0.005	−0.022	−0.003	−0.004
(ref: non-health worker)	(0.011)	(0.019)	(0.013)	(0.020)	(0.015)	(0.015)
Region2	0.000	−0.052**	−0.001	0.030*	0.023	0.000
	(0.015)	(0.021)	(0.016)	(0.019)	(0.016)	(0.021)
Region3	0.002	0.013	−0.001	−0.014	−0.018	0.019
	(0.011)	(0.016)	(0.012)	(0.011)	(0.013)	(0.015)
Region4	0.007	0.009	−0.006	0.002	−0.023*	0.018
	(0.016)	(0.018)	(0.023)	(0.015)	(0.018)	(0.016)
Region5	0.006	0.039*	−0.011	0.000	−0.042**	0.009
	(0.016)	(0.023)	(0.015)	(0.018)	(0.021)	(0.022)
Region 6	0.068***	0.030	−0.047***	0.004	−0.038**	−0.018
	(0.025)	(0.033)	(0.017)	(0.020)	(0.018)	(0.019)
Region 7	0.015	0.004	0.048*	−0.015	−0.039*	−0.014
	(0.020)	(0.034)	(0.025)	(0.021)	(0.021)	(0.021)
Region 8	−0.003	0.041**	0.008	−0.011	−0.021	−0.015
	(0.014)	(0.021)	(0.016)	(0.016)	(0.017)	(0.019)
Region 9	0.008	0.034*	0.019*	−0.010	−0.027	−0.024
	(0.013)	(0.019)	(0.011)	(0.013)	(0.018)	(0.017)
Region 10	0.003	0.002	0.028	−0.003	−0.010	−0.021
	(0.011)	(0.017)	(0.018)	(0.012)	(0.014)	(0.017)
Observations	322	322	322	322	322	322

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Note: Separate regressions were run to accommodate a change in reference category for the variables 'income', 'health usage' and 'employment'. The results are combined in the above table.

Table 7.7: Average partial effects using a FML model (females)

Demographic characteristics	Only available treatment	Need	Age	Societal benefit	Individual benefit	Lifestyle
European/Maori, Maori	−0.013	0.052*	−0.031*	0.006	−0.002	−0.013
(ref: European)	(0.016)	(0.029)	(0.018)	(0.025)	(0.019)	(0.022)
Other ethnicity	−0.003	0.017	−0.013	−0.002	−0.018	0.020
(ref: European)	(0.014)	(0.026)	(0.018)	(0.020)	(0.021)	(0.021)
Age (continuous)	0.002	−0.006	**−0.011****	0.007	**0.012*****	−0.005
Mean age 45-54 years	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)
Income $30,001-$70,000	−0.017	0.006	0.008	−0.010	0.027	−0.014
(ref: $0-$30,000)	(0.013)	(0.021)	(0.017)	(0.013)	(0.019)	(0.016)
Income over $70,000	**−0.031****	−0.013	0.011	0.002	**0.054*****	−0.022
(ref: $0-$30,000)	(0.015)	(0.021)	(0.017)	(0.014)	(0.020)	(0.018)
Income over $70,000	−0.015	−0.018	0.003	0.012	0.026*	−0.008
(ref: $30,001-$70,000)	(0.011)	(0.017)	(0.015)	(0.014)	(0.016)	(0.015)
Not living with children	−0.004	0.008	0.007	−0.006	−0.007	0.002
(ref: living with children)	(0.010)	(0.015)	(0.011)	(0.012)	(0.012)	(0.013)
Occasional health use	0.001	−0.024	**−0.028****	0.000	0.030*	0.022
(ref: seldom health use)	(0.011)	(0.018)	(0.013)	(0.012)	(0.016)	(0.020)
Frequent health use	−0.014	0.023	−0.020	0.000	0.004	0.017
(ref: seldom health use)	(0.013)	(0.017)	(0.013)	(0.016)	(0.017)	(0.019)
Frequent health use	−0.015	**0.038****	0.008	0.000	−0.026*	−0.006
(ref: occasional health use)	(0.011)	(0.017)	(0.014)	(0.012)	(0.015)	(0.015)
No serious illness	0.011	−0.030	0.013	0.008	−0.022	0.020
(ref: serious illness)	(0.025)	(0.027)	(0.025)	(0.019)	(0.025)	(0.045)
No health insurance	0.000	0.001	0.002	0.007	−0.007	−0.003
(ref: health insurance)	(0.009)	(0.014)	(0.011)	(0.013)	(0.012)	(0.014)
Sec school qualifications	−0.014	−0.010	0.016	0.021	−0.015	0.002
(ref: no qualifications)	(0.024)	(0.027)	(0.033)	(0.024)	(0.022)	(0.038)
Other post sec school quals	−0.009	−0.014	−0.004	0.028	0.005	−0.006
(ref: no qualifications)	(0.025)	(0.027)	(0.032)	(0.024)	(0.022)	(0.037)
Degree or equivalent	0.012	−0.019	−0.003	0.032	−0.008	−0.015
(ref: no qualifications)	(0.027)	(0.027)	(0.032)	(0.026)	(0.023)	(0.037)
Not working	0.028*	0.017	−0.017	−0.009	0.014	**−0.033****
(ref: working)	(0.014)	(0.016)	(0.013)	(0.012)	(0.014)	(0.015)
Retired	0.004	−0.022	0.025	−0.018	0.022	−0.011
(ref: working)	(0.015)	(0.022)	(0.020)	(0.013)	(0.018)	(0.016)
Retired	−0.021	−0.040*	0.043	−0.011	0.005	0.024
(ref: not working)	(0.016)	(0.024)	(0.027)	(0.017)	(0.022)	(0.025)
Health worker	−0.009	0.040	−0.004	−0.027	−0.007	0.007
(ref: non-health worker)	(0.016)	(0.025)	(0.014)	(0.026)	(0.018)	(0.020)
Region2	0.007	-0.070**	0.004	0.056*	0.043*	-0.040
	(0.023)	(0.028)	(0.029)	(0.030)	(0.023)	(0.031)
Region3	0.005	-0.007	-0.009	0.006	-0.024	0.029
	(0.015)	(0.018)	(0.014)	(0.015)	(0.016)	(0.020)
Region4	-0.014	0.000	0.016	0.002	-0.030*	0.025
	(0.017)	(0.023)	(0.030)	(0.020)	(0.016)	(0.022)
Region5	0.023	0.039	-0.015	0.005	-0.047	-0.006
	(0.020)	(0.030)	(0.019)	(0.025)	(0.029)	(0.027)
Region 6	0.069**	-0.038	-0.027	0.025	-0.030	0.001
	(0.032)	(0.041)	(0.023)	(0.029)	(0.018)	(0.026)
Region 7	0.031	0.009	0.052	-0.017	-0.051*	-0.024
	(0.027)	(0.040)	(0.034)	(0.026)	(0.027)	(0.026)
Region 8	-0.006	0.031	0.012	-0.002	-0.020	-0.014
	(0.018)	(0.030)	(0.019)	(0.019)	(0.021)	(0.027)
Region 9	0.008	0.020	0.032**	-0.004	-0.033*	-0.022
	(0.016)	(0.023)	(0.015)	(0.016)	(0.019)	(0.019)
Region 10	0.011	-0.005	0.030	0.015	-0.036*	-0.015
	(0.016)	(0.022)	(0.024)	(0.018)	(0.019)	(0.023)
Observations	322	322	322	322	322	322

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Note: Separate regressions were run to accommodate a change in reference category for the variables 'income', 'health usage' and 'employment'. The results are combined in the above table.

Table 7.8: Average partial effects using a FML model (males)

Demographic characteristics	Only available treatment	Need	Age	Societal benefit	Individual benefit	Lifestyle
European/Maori, Maori	−0.009	0.042	−0.017	−0.008	0.006	−0.014
(ref: European)	(0.016)	(0.026)	(0.020)	(0.017)	(0.024)	(0.018)
Other ethnicity	0.056	0.018	−0.002	−0.045*	0.020	**−0.047***
(ref: European)	(0.036)	(0.030)	(0.034)	(0.026)	(0.040)	(0.027)
Age (continuous)	−0.002	0.005	−0.010*	−0.005	0.002	0.010
Mean age 45-54 years	(0.005)	(0.007)	(0.006)	(0.005)	(0.006)	(0.007)
Income $30,001-$70,000	−0.012	−0.022	−0.016	0.000	0.021	0.030
(ref: $0−$30,000)	(0.017)	(0.028)	(0.019)	(0.022)	(0.026)	(0.032)
Income over $70,000	−0.011	**−0.071****	−0.034	0.021	0.028	0.066
(ref: $0-$30,000)	(0.016)	(0.029)	(0.023)	(0.026)	(0.028)	(0.046)
Income over $70,000	0.001	**−0.049*****	−0.018	0.023	0.008	0.035
(ref: $30,001-$70,000)	(0.016)	(0.019)	(0.016)	(0.018)	(0.019)	(0.026)
Not living with children	0.017	−0.002	−0.008	0.011	−0.020	0.003
(ref: living with children)	(0.015)	(0.018)	(0.018)	(0.014)	(0.017)	(0.014)
Occasional health use	−0.022	0.000	0.000	−0.001	0.023	0.001
(ref: seldom health use)	(0.022)	(0.021)	(0.018)	(0.015)	(0.019)	(0.017)
Frequent health use	−0.011	0.014	0.015	−0.003	0.014	−0.030
(ref: seldom health use)	(0.015)	(0.025)	(0.024)	(0.019)	(0.024)	(0.026)
Frequent health use	0.012	−0.001	0.015	−0.002	−0.008	−0.030
(ref: occasional health use)	(0.015)	(0.020)	(0.020)	(0.018)	(0.019)	(0.020)
No serious illness	−0.022	−0.004	0.031	0.000	−0.003	0.000
(ref: serious illness)	(0.049)	(0.032)	(0.057)	(0.016)	(0.026)	(0.017)
No health insurance	0.007	−0.011	0.014	−0.010	0.000	0.001
(ref: health insurance)	(0.017)	(0.020)	(0.026)	(0.019)	(0.016)	(0.014)
Sec school qualifications	0.042	0.006	0.017	0.001	−0.020	−0.046*
(ref: no qualifications)	(0.033)	(0.028)	(0.028)	(0.020)	(0.034)	(0.024)
Other post sec school quals	0.012	0.031	0.017	−0.003	−0.013	−0.044*
(ref: no qualifications)	(0.023)	(0.031)	(0.029)	(0.020)	(0.032)	(0.025)
Degree or equivalent	0.025	0.025	−0.002	−0.007	0.008	−0.049*
(ref: no qualifications)	(0.024)	(0.030)	(0.027)	(0.020)	(0.033)	(0.027)
Not working	−0.022	0.003	−0.035*	0.000	**0.055****	−0.001
(ref: working)	(0.020)	(0.027)	(0.020)	(0.020)	(0.024)	(0.029)
Retired	0.007	−0.020	−0.013	0.016	0.011	−0.001
(ref: working)	(0.021)	(0.029)	(0.018)	(0.023)	(0.023)	(0.024)
Retired	0.031	−0.027	0.025	0.014	−0.041	−0.002
(ref: not working)	(0.036)	(0.032)	(0.028)	(0.025)	(0.028)	(0.034)
Health worker	0.003	0.016	−0.028	0.031	0.037	−0.059
(ref: non-health worker)	(0.035)	(0.040)	(0.046)	(0.039)	(0.028)	(0.066)
Region2	0.002	−0.044	−0.002	0.008	0.017	0.019
	(0.019)	(0.030)	(0.019)	(0.018)	(0.023)	(0.026)
Region3	−0.020	0.042	0.019	−0.041*	−0.005	0.005
	(0.016)	(0.031)	(0.020)	(0.022)	(0.024)	(0.022)
Region4	0.029	0.016	−0.033	0.005	−0.024	0.007
	(0.028)	(0.030)	(0.040)	(0.023)	(0.033)	(0.022)
Region5	−0.006	−0.020	0.009	0.010	−0.029	0.035
	(0.029)	(0.032)	(0.024)	(0.027)	(0.022)	(0.044)
Region 6	0.056*	0.124***	−0.071**	−0.030	−0.037	−0.042
	(0.034)	(0.046)	(0.032)	(0.030)	(0.033)	(0.030)
Region 7	−0.010	−0.046	0.037	0.007	−0.007	0.019
	(0.024)	(0.038)	(0.042)	(0.028)	(0.025)	(0.020)
Region 8	−0.010	0.062**	0.002	−0.025	−0.034	0.005
	(0.017)	(0.028)	(0.023)	(0.027)	(0.026)	(0.023)
Region 9	0.019	0.059	−0.019	−0.036	0.002	−0.025
	(0.029)	(0.037)	(0.022)	(0.026)	(0.047)	(0.033)
Region 10	−0.006	−0.008	0.023	−0.018	0.032	−0.023
	(0.016)	(0.023)	(0.026)	(0.019)	(0.020)	(0.028)
Observations	322	322	322	322	322	322

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Note: Separate regressions were run to accommodate a change in reference category for the variables ‘income’, ‘health usage’ and ‘employment’. The results are combined in the above table.

As discussed in Section 7.5, because of the low response rate and a potential non-coverage bias, a conservative approach is taken in interpreting the significant results – only the demographic characteristics that have coefficients which are statistically significant at the 5% or 1% level will be discussed.[158]

7.7.1 Ethnicity

There is a statistically significant effect at the 5% level between Maori (*n*=27) and non-Maori (*n*=275) respondents on the criterion 'need'. Respondents who identify as Maori, on average, place more importance (+4.4%) on 'need' (relative to the other criteria) compared with non-Maori respondents.

As previously mentioned, Maori have comparatively higher health needs than non-Maori (Ministry of Health 2011) which is perhaps why Maori place comparatively more importance on 'need'.

7.7.2 Age of respondents

Respondents older than 54, on average, place *less* importance on the criterion 'age' (–1.1% at $p < 0.001$ relative to the other criteria) and more importance on 'individual benefit' (+0.8% at $p < 0.05$ relative to the other criteria) compared to respondents younger than 45. It appears that female respondents are influencing these results. Female respondents over 54 years place *less* importance on 'age' relative to the other criteria (–1.1% $p < 0.05$) and *more* importance on 'individual benefit' relative to the other criteria (+1.2%, $p < 0.001$) compared to female respondents younger than 45.

This means that younger respondents show a greater preference for treating younger members of the population but less preference for individual benefit compared to older respondents.

7.7.3 Household income

On average, respondents with a household income over $70,000 place *les*s importance on 'need' relative to the other criteria (–3.1%, $p < 0.05$) and *more* on importance on 'individual

[158] Also, as explained in Section 7.5 the statistically significant 'region' coefficients will not be discussed as interpretation of the statistically significant coefficients is of little value to the analysis.

benefit' relative to the other criteria (+3.4%, $p < 0.05$) compared to respondents with a household income less than $30,000. When analysing the results by gender, it appears that male respondents are influencing the statistically significant result for 'need' relative to the other criteria (–7.1% at $p < 0.05$) and that female respondents are influencing the statistically significant effect for 'individual benefit' relative to the other criteria (+5.4 %, $p < 0.001$). In addition, on average, female respondents with a household income over $70,000 place less importance on 'only available treatment' relative to the other criteria (–3.1%, $p < 0.05$) compared to females with a household income less than $30,000.

Respondents with a household income greater than $70,000, on average, place less importance on 'need' relative to the other criteria (–2.8%, $p < 0.05$) compared to respondents with a household income of $30,001-$70.000. When analysing the results by gender, male respondents in the higher income group place less importance on 'need' relative to the other criteria (–4.9%, $p < 0.001$) compared to the males in the middle income group.

High-income respondents (particularly male respondents) place relatively *less* importance on 'need' compared to middle and low-income respondents. On the other hand, middle-income respondents (particularly female respondents) place relatively *more* importance on 'individual benefit' compared to high-income respondents.

7.7.4 Health usage

On average, respondents who use health services occasionally, place *more* importance on 'individual benefit' relative to the other criteria (+2.9%, $p < 0.05$) and *less* importance on 'age' relative to the other criteria (–2.8%, $p < 0.05$) compared to respondents who seldom use health services. In particular, female respondents who use health services occasionally, place *less* importance on treating the younger members of the population (–2.28%, $p < 0.05$) compared to females who seldom use health services.

In addition, females who use health services occasionally, on average, place *less* importance on 'need' relative to the other criteria (–3.8%, $p < 0.05$) compared to females who use health services frequently.

7.7.5 Type of worker

There is a statistically significant effect between respondents who work in the health field (n=33) and respondents who do not (n=289) on the criterion 'need'. Respondents who are health care workers, on average, place more importance on 'need' relative to the other criteria (+3.6%, $p < 0.05$) compared to respondents who are not health care workers.

7.7.6 Work status

There are no statistically significant effects at the 5% or 1% levels when analysing the entire sample. However, when the male and female respondents are considered separately, female respondents who do *not* work place less importance on the criterion 'lifestyle' relative to other criteria (–3.3%, $p < 0.05$) compared to female respondents who work and male respondents who do *not* work place more importance on 'individual benefit' relative to other criteria (+5.5%, $P < 0.05$) compared to males respondents who work.

7.7.7 Discussion of FML regression results

The results of the FML regressions reveal that several of the demographic characteristics have a statistically significant effect on the criteria weights. For example, females who use health services frequently place relatively more importance on 'need' compared to females who use health services occasionally. This seems reasonable as it could be assumed that one reason for seeking health services frequently is that a person is in greater need and would therefore place more importance on need. Similarly, health care workers place more importance on 'need' compared to respondents who do not work in health care. Once again, this seems reasonable as given their occupation, health care workers might place more importance on 'need' compared to respondents who are not health workers. Interestingly, but perhaps not surprisingly, is that younger respondents show a greater preference for treating younger members of the population compared to older respondents.

In terms of other demographic characteristics, Maori respondents place relatively *more* importance on 'need' compared to non-Maori, and middle and low-income respondents place relatively *more* importance on 'need' compared to high-income respondents. These results support the discussion in Section 7.3; Maori have comparatively higher health needs than non-Maori (Ministry of Health 2011) and therefore it seems reasonable that they place more

importance on 'need' compared to non-Maori. Similarly, people on low incomes are more likely to have lower life expectancies and higher health needs compared to people on middle or high incomes (Swartz 2009) and therefore it seems reasonable that lower-income respondents consider 'need' to be more important compared to respondents on high incomes.

7.8 Regression diagnostics

To be able to generalise the results from a regression model to the wider population several underlying assumptions need to be met (Field 2009). These are discussed below.

7.8.1 Normally distributed errors

With OLS the requirement is that the errors[159] are identically and independently distributed. These two assumptions are discussed below (homoscedasticity and independence).

7.8.2 Homoscedasticity

For the assumption of homoscedasticity to hold, the variance of the residuals ought to be constant. This means that the variance of a criterion's weights should be equal at all levels of a demographic characteristic. The Breusch-Pagan test is used to test whether the variance of the residuals are dependent on the values of the independent variables (i.e. the demographic characteristics).

Table 7.9: Breusch-Pagan test for homoscedasticity

Criteria (*N*=322)	Chi2(1)	Prob > chi2
Only available treatment	9.08	0.003
Age	17.36	0.000
Societal benefit	5.95	0.015
Need	2.12	0.146
Individual benefit	0.42	0.515
Lifestyle	1.26	0.262

[159] 'Errors', also known as 'residuals', are the differences between the observed values and the predicted values.

As can be seen in Table 7.9, $p < 0.5$ (indicating that the residuals are homogenous) for only three of the criteria: 'need', 'individual benefit' and 'lifestyle'. However, this test is sensitive to other assumptions such as normality of the data. When the data are tested for normal distribution, only two of the criteria are normally distributed ('need' and 'individual benefit'). However, in large samples variables indicating statistically significant skewness often do not diverge enough from normality to make a significant difference to the analysis (Tabachnick & Fidell 2007). Similarly, the larger the sample size, the less the impact kurtosis has on the analysis. According to Waternaux (1976), estimates of variance associated with positive kurtosis become more accurate with sample size greater than 100; and with negative kurtosis, with sample size greater than 200.

The Breusch-Pagan test is therefore used in combination with scatter plots to assess homoscedasticity. Though the residuals for 'only available treatment', 'age' and 'societal benefit' do not exhibit homogeneity, when the residuals are plotted against the fitted values (or predicted values) the plots are not markedly different compared to the other three criteria (see Appendix 7.3).

7.8.3 Independence

Independence of the residuals requires that the residuals associated with one respondent's set of weights are not correlated with the residuals of any other set of weights. The degree of multicollinearity between the demographic characteristics can be measured using the variance inflation factor (VIF). The 'tolerance value' is equal to 1/VIF. As a rule of thumb a tolerance value less than 0.1 (equal to a VIF over 10) indicates multicollinearity.

Table 7.10: Variance inflation factor for the demographic characteristics

Demographic characteristics	VIF	1/VIF
Gender	1.11	0.9013
Ethnicity1	1.15	0.8719
Ethnicity 2	1.18	0.8111
Age	1.90	0.5262
Income1	2.19	0.4572
Income2	2.50	0.4003
Income3	1.71	0.5852
Household composition	1.35	0.7400
Healthuse1	1.50	0.6675
Healthuse2	1.62	0.6173
Serious illness	1.14	0.8738
Health insurance	1.19	0.8431
Qualifications1	4.82	0.2074
Qualifications2	4.35	0.2298
Qualifications3	5.05	0.1980
Employment1	1.41	0.7093
Employment2	1.76	0.5678
Worker type	1.20	0.8337
Region1	1.22	0.8211
Region2	1.43	0.7008
Region3	1.23	0.8147
Region4	1.23	0.8105
Region5	1.20	0.8328
Region6	1.19	0.8392
Region7	1.27	0.7875
Region8	1.30	0.7678
Region9	1.36	0.7354
Mean VIF	**1.80**	

When a variable has a tolerance value close to 0.1 it is possible that the variable is a linear combination of some of the other variables. As can be seen in Table 7.10, the lowest tolerance values are 0.2074, 0.2298 and 0.1980 for the demographic characteristic 'qualifications'. When the variable 'qualifications' is deleted from the demographic groups the mean VIF changes to 1.41. As deleting 'qualifications' makes little difference to the VIF, it is kept in the analysis.[160]

[160] Two OLS regressions were run with 'qualifications' included in one of the regressions and omitted in the other. Omitting 'qualifications' had no affect on the results.

7.8.4 Linearity

The Ramsey Regression Equation Specification Error Test (RESET) is a general specification test for a linear regression model (Ramsey 1969). The RESET tests the null hypothesis that non-linear combinations of the fitted values help explain the dependent variable (i.e. the respondents' criteria weights). If the null hypothesis is true it means that *non-linear* combinations of the demographic characteristics affect the criteria weights and therefore the model is incorrectly specified.

Table 7.11: RESET results

Criteria (*N*=322)	F(3,291)	Prob > F
Only available treatment	1.26	0.287
Age	0.71	0.548
Societal benefit	1.18	0.318
Need	0.27	0.846
Individual benefit	1.22	0.302
Lifestyle	0.85	0.466

As can be seen in Table 7.11, $p > 0.05$ for all the criteria. Therefore the null hypothesis can be rejected.

7.8.5 Statistical significance

Looking back at Tables 7.6-7.8, it can be seen that approximately 5% of the coefficients are statistically significant at the 5% level of confidence. To determine whether the statistically significant results are by chance only, I split the random sample data into two samples and repeated the regression analyses. I found that the statistically significant results are split between the samples, i.e., demographic characteristics that are statistically significant in one sample are *not* statistically significant in the other. This indicates that the statistically significant results may be by chance only. Therefore, though several of the demographic characteristics have an effect on the criteria weights (in a way that might be expected) it appears possible that the variation in preferences is much less related to particular

demographic characteristics than to respondents' personalities and lifestyles or to other factors not captured in this survey.

7.9 Conclusion

Each respondent from the random sample has a utility function which represents their preferences for the six criteria. In this chapter regression analysis is used to determine whether the demographic characteristics of respondents can predict respondents' preferences for the criteria. As utility between respondents is not directly comparable, the MRS between pairs of criteria are regressed against the demographic characteristics of the respondents. Several statistically significant effects are found.

OLS, SUR and the FML model are also used to explore the relationship between the demographic characteristics of the respondents and the criteria weights. The results of these regressions are similar, producing several statistically significant effects. However, the results should be treated with caution given the 10% survey response rate and possible non-coverage bias. Also when the random sample is split, the statistically significant results are not common to both samples suggesting that the statistically significant results may be by chance only. It appears possible that the variation in respondents' preferences is largely idiosyncratic and not directly related to the respondents' characteristics.

To gain a better understanding of the heterogeneity of respondents' preferences, I also used cluster analysis – a data reduction method – which is discussed in the next chapter.

Appendix 7.1: Results of OLS

Demographic characteristics	Only available treatment	Need	Age	Societal benefit	Individual benefit	Lifestyle
Male	−1.054	1.554	0.143	0.409	0.198	−1.235
(ref: female)	(0.726)	(1.055)	(0.910)	(0.781)	(0.918)	(0.945)
European/Maori, Maori	−0.0938	**4.198****	**−2.528***	−0.223	−0.339	−1.032
(ref: European)	(1.310)	(1.811)	(1.296)	(1.428)	(1.594)	(1.473)
Other ethnicity	0.981	3.090	−0.115	−2.322	−1.211	−0.421
(ref: European)	(1.334)	(2.096)	(1.884)	(1.577)	(1.942)	(1.813)
Age (continuous)	0.284	−0.170	**−1.021*****	0.0639	**0.814****	0.0305
Mean age 45-54 years	(0.332)	(0.451)	(0.392)	(0.355)	(0.379)	(0.344)
Income $30,001−$70,000	−1.396	−0.192	0.428	−0.465	1.788	−0.142
(ref: $0−$30,000)	(1.205)	(1.556)	(1.305)	(1.081)	(1.406)	(1.408)
Income over $70,000	−2.189	**−2.931***	0.293	0.929	**3.286****	0.615
(ref: $0−$30,000)	(1.342)	(1.601)	(1.329)	(1.164)	(1.395)	(1.485)
Income over $70,000	−0.793	**−2.739****	−0.135	1.394	1.498	0.757
(ref: $30,001−$70,000)	(0.938)	(1.338)	(1.110)	(0.983)	(1.151)	(1.169)
Not living with children	0.661	0.874	−0.275	−0.110	−1.610	0.475
(ref: living with children)	(0.785)	(1.167)	(0.956)	(0.894)	(1.055)	(1.033)
Occasional health use	−0.851	−1.565	**−2.016****	−0.118	**2.819*****	1.736
(ref: seldom health use)	(0.817)	(1.275)	(0.988)	(0.927)	(1.055)	(1.075)
Frequent health use	−0.669	0.697	−1.048	0.531	0.942	−0.446
(ref: seldom health use)	(1.089)	(1.549)	(1.288)	(1.314)	(1.426)	(1.412)
Frequent health use	0.182	2.262	0.969	0.650	−1.878	**−2.182***
(ref: occasional health use)	(0.922)	(1.372)	(1.064)	(1.113)	(1.246)	(1.164)
No serious illness	−0.0915	**−2.379****	**1.797****	0.650	−0.933	0.968
(ref: serious illness)	(0.740)	(1.183)	(0.910)	(0.840)	(0.961)	(0.955)
No health insurance	0.433	−0.918	0.378	0.234	−0.604	0.482
(ref: health insurance)	(0.741)	(1.094)	(0.878)	(0.805)	(0.878)	(0.896)
Sec school qualifications	−0.461	−0.0122	1.907	0.712	−1.614	−0.547
(ref: no qualifications)	(1.794)	(1.996)	(2.169)	(1.449)	(2.038)	(2.419)
Other post sec school quals	−0.400	0.631	0.323	1.247	−0.627	−1.184
(ref: no qualifications)	(1.856)	(2.082)	(2.229)	(1.464)	(2.056)	(2.437)
Degree or equivalent	0.703	0.0186	0.0241	1.184	−0.267	−1.679
(ref: no qualifications)	(1.797)	(2.024)	(2.192)	(1.467)	(2.028)	(2.437)
Not working	1.288	1.041	**−2.006***	−0.388	**2.265***	−2.205
(ref: working)	(1.059)	(1.449)	(1.123)	(1.050)	(1.230)	(1.340)
Retired	0.592	−1.592	0.926	−0.345	0.351	0.0782
(ref: working)	(1.251)	(1.739)	(1.273)	(1.120)	(1.377)	(1.425)
Retired	−0.696	−2.633	**2.932***	0.0433	−1.914	2.283
(ref: not working)	(1.576)	(2.045)	(1.649)	(1.447)	(1.793)	(1.764)
Health worker	−0.175	**3.553****	−0.461	−2.146	−0.384	−0.377
(ref: non−health worker)	(1.091)	(1.609)	(1.257)	(1.547)	(1.422)	(1.546)
	(1.483)	(2.067)	(1.563)	(1.832)	(1.717)	(2.204)
Region 2	−0.142	−5.049**	−0.106	2.982	2.441	−0.104
	(1.483)	(2.067)	(1.563)	(1.832)	(1.717)	(2.204)
Region 3	0.168	1.269	−0.111	−1.400	−1.915	1.968
	(1.069)	(1.611)	(1.222)	(1.164)	(1.443)	(1.515)
Region 4	0.639	0.882	−0.495	0.196	−3.017	1.795
	(1.587)	(1.800)	(2.322)	(1.507)	(1.908)	(1.711)
Region 5	0.523	3.899	−0.981	0.0500	−4.334*	0.855
	(1.603)	(2.383)	(1.472)	(1.934)	(2.207)	(2.260)
Region 6	6.461***	3.199	−4.744***	0.583	−3.663**	−1.825
	(1.996)	(3.357)	(1.616)	(2.028)	(1.851)	(1.980)
Region 7	1.515	0.650	4.469*	−1.377	−4.019*	−1.232
	(1.986)	(3.449)	(2.391)	(2.194)	(2.292)	(2.102)
Region 8	−0.185	4.092*	0.850	−1.075	−2.180	−1.498
	(1.393)	(2.142)	(1.684)	(1.610)	(1.772)	(1.870)

Region 9	0.823	3.389*	1.963*	−0.986	−2.681	−2.495
	(1.326)	(1.874)	(1.135)	(1.409)	(1.863)	(1.686)
Region 10	0.317	0.201	2.742	−0.246	−1.031	−1.990
	(1.144)	(1.718)	(1.704)	(1.296)	(1.543)	(1.645)
Constant OLS1	11.37***	28.75***	15.54***	11.04***	20.28***	12.95***
	(3.200)	(4.343)	(3.721)	(3.554)	(3.905)	(4.371)
Constant OLS2	10.42***	28.04***	11.94***	10.07***	27.15***	12.34***
	(3.549)	(4.559)	(3.636)	(3.378)	(3.857)	(4.400)
Observations	322	322	322	322	322	322
R−squared	0.096	0.161	0.124	0.062	0.147	0.081

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Note: Separate regressions were run to accommodate a change in reference category for the variables ‘income’, ‘health usage’ and ‘employment’. The results are combined in the above table.

Appendix 7.2: Results of OLS (SUR)

Demographic characteristics	Only available treatment	Need	Age	Societal benefit	Individual benefit	Lifestyle
Male	−1.054	1.554	0.143	0.409	0.198	−1.235
(ref: female)	(0.695)	(1.010)	(0.871)	(0.747)	(0.878)	(0.905)
European/Maori, Maori	−0.0938	4.198**	−2.528**	−0.223	−0.339	−1.032
(ref: European)	(1.254)	(1.733)	(1.240)	(1.367)	(1.525)	(1.410)
Other ethnicity	0.981	3.090	−0.115	−2.322	−1.211	−0.421
(ref: European)	(1.277)	(2.005)	(1.803)	(1.509)	(1.859)	(1.735)
Age (continuous)	0.284	−0.170	−1.021***	0.0639	0.814**	0.0305
Mean age 45-54 years	(0.318)	(0.432)	(0.375)	(0.340)	(0.363)	(0.329)
Income $30,001−$70,000	−1.396	−0.192	0.428	−0.465	1.788	−0.142
(ref: $0−$30,000)	(1.153)	(1.489)	(1.249)	(1.034)	(1.345)	(1.348)
Income over $70,000	−2.189*	−2.931*	0.293	0.929	3.286**	0.615
(ref: $0−$30,000)	(1.284)	(1.532)	(1.272)	(1.114)	(1.335)	(1.421)
Income over $70,000	−0.793	−2.739**	−0.135	1.394	1.498	0.757
(ref: $30,001−$70,000)	(0.898)	(1.280)	(1.062)	(0.941)	(1.101)	(1.119)
Not living with children	0.661	0.874	−0.275	−0.110	−1.610	0.475
(ref: living with children)	(0.751)	(1.117)	(0.915)	(0.856)	(1.010)	(0.988)
Occasional health use	−0.851	−1.565	−2.016**	−0.118	2.819***	1.736*
(ref: seldom health use)	(0.782)	(1.220)	(0.945)	(0.887)	(1.010)	(1.029)
Frequent health use	−0.669	0.697	−1.048	0.531	0.942	−0.446
(ref: seldom health use)	(1.042)	(1.482)	(1.232)	(1.258)	(1.365)	(1.351)
Frequent health use	0.182	2.262*	0.969	0.650	−1.878	−2.182*
(ref: occasional health use)	(0.883)	(1.314)	(1.018)	(1.066)	(1.193)	(1.114)
No serious illness	−0.0915	−2.379**	1.797**	0.650	−0.933	0.968
(ref: serious illness)	(0.708)	(1.132)	(0.871)	(0.804)	(0.920)	(0.914)
No health insurance	0.433	−0.918	0.378	0.234	−0.604	0.482
(ref: health insurance)	(0.709)	(1.047)	(0.840)	(0.770)	(0.840)	(0.857)
Sec school qualifications	−0.461	−0.0122	1.907	0.712	−1.614	−0.547
(ref: no qualifications)	(1.717)	(1.910)	(2.075)	(1.387)	(1.950)	(2.315)
Other post sec school quals	−0.400	0.631	0.323	1.247	−0.627	−1.184
(ref: no qualifications)	(1.776)	(1.992)	(2.133)	(1.401)	(1.968)	(2.332)
Degree or equivalent	0.703	0.0186	0.0241	1.184	−0.267	−1.679
(ref: no qualifications)	(1.720)	(1.937)	(2.098)	(1.404)	(1.941)	(2.332)
Not working	1.288	1.041	−2.006*	−0.388	2.265*	−2.205*
(ref: working)	(1.014)	(1.386)	(1.074)	(1.005)	(1.178)	(1.283)
Retired	0.592	−1.592	0.926	−0.345	0.351	0.0782
(ref: working)	(1.197)	(1.664)	(1.219)	(1.072)	(1.318)	(1.364)
Retired	−0.696	−2.633	2.932*	0.0433	−1.914	2.283
(ref: not working)	(1.508)	(1.957)	(1.578)	(1.385)	(1.715)	(1.688)
Health worker	−0.175	3.553**	−0.461	−2.146	−0.384	−0.377
(ref: non−health worker)	(1.044)	(1.540)	(1.203)	(1.481)	(1.361)	(1.480)
Region 2	−0.142	−5.049**	−0.106	2.982*	2.441	−0.104
	(1.420)	(1.978)	(1.496)	(1.754)	(1.643)	(2.109)
Region 3	0.168	1.269	−0.111	−1.400	−1.915	1.968
	(1.023)	(1.541)	(1.169)	(1.114)	(1.381)	(1.450)
Region 4	0.639	0.882	−0.495	0.196	−3.017*	1.795
	(1.518)	(1.722)	(2.223)	(1.442)	(1.826)	(1.638)
Region 5	0.523	3.899*	−0.981	0.0500	−4.334**	0.855
	(1.534)	(2.280)	(1.408)	(1.850)	(2.112)	(2.163)
Region 6	6.461***	3.199	−4.744***	0.583	−3.663**	−1.825
	(1.910)	(3.213)	(1.547)	(1.941)	(1.771)	(1.895)
Region 7	1.515	0.650	4.469*	−1.377	−4.019*	−1.232
	(1.901)	(3.301)	(2.288)	(2.100)	(2.194)	(2.012)
Region 8	−0.185	4.092**	0.850	−1.075	−2.180	−1.498
	(1.333)	(2.050)	(1.612)	(1.541)	(1.696)	(1.789)

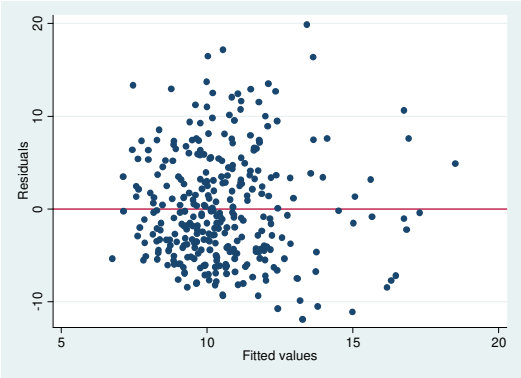
Region 9	0.823	3.389*	1.963*	−0.986	−2.681	−2.495
	(1.269)	(1.794)	(1.086)	(1.349)	(1.783)	(1.614)
Region 10	0.317	0.201	2.742*	−0.246	−1.031	−1.990
	(1.094)	(1.644)	(1.630)	(1.241)	(1.477)	(1.574)
Constant	11.37***	28.75***	15.54***	11.04***	20.28***	12.95***
	(3.062)	(4.156)	(3.561)	(3.401)	(3.737)	(4.183)
Constant	10.42***	28.04***	11.94***	10.07***	27.15***	12.34***
	(3.397)	(4.363)	(3.480)	(3.232)	(3.691)	(4.211)
Observations	322	322	322	322	322	322

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

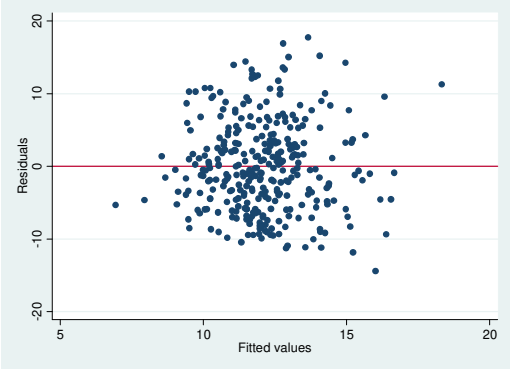
Note: Separate regressions were run to accommodate a change in reference category for the variables ‘income’, ‘health usage’ and ‘employment’. The results are combined in the above table.

Appendix 7.3: Scatter plots to test for homoscedasticity

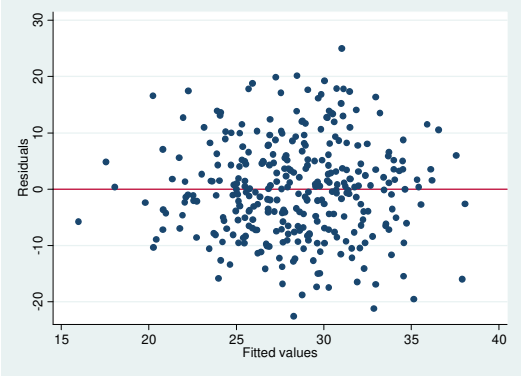
Only available treatment



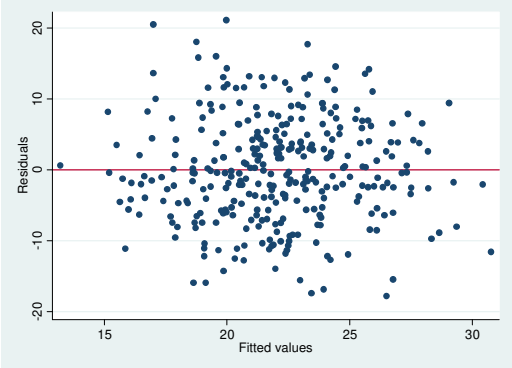
Societal benefit



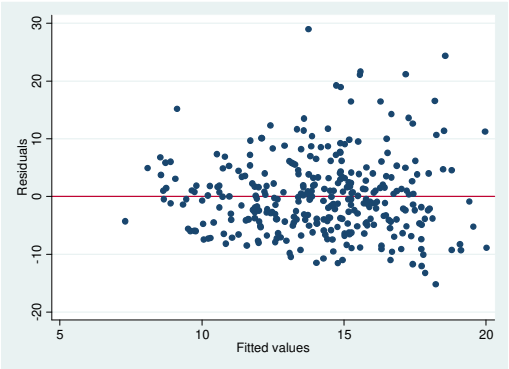
Need



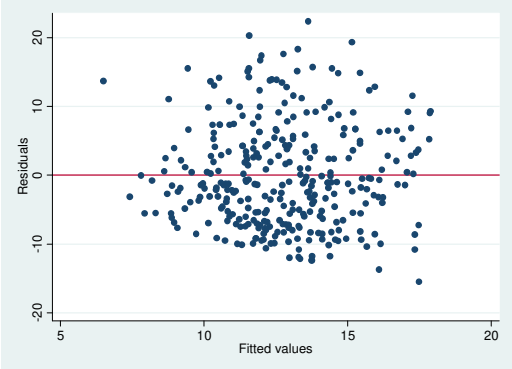
Individual benefit



Age



Lifestyle



~ Chapter 8 ~

Cluster analysis

8.1 Introduction

Cluster analysis is a statistical method for identifying natural groupings in data. It is used in this thesis to 'reduce' the heterogeneity of the data. From a policy perspective, the main objective of using cluster analysis is to gain a better understanding of the patterns of preferences of respondents.

In this chapter the technical process of forming clusters (groups) from the survey data is explained, and the results of the cluster analysis are presented and discussed. The mean criteria weights from the clusters are then used to rank health vignettes to illustrate how respondents' preferences affect the relative rankings of health treatments.

8.2 Cluster analysis

Each random sample respondent has their own set of criteria weights which represent the relative importance of the six criteria to that respondent. Each respondent's set of weights remains together in the clustering process and is referred to as a case. To form clusters a hierarchical or non-hierarchical procedure can be used (Norusis 2009).

In non-hierarchical clustering, such as k-means clustering, a clustering algorithm starts with an initial set of means criteria weights or 'seeds' and then classifies the cases based on their distances to the centre of these clusters (distance measures will be discussed in the next section). The algorithm repeatedly reassigns cases to clusters so that a case can move from one cluster to another. The number of clusters (k) needs to be predetermined. If the clusters do not produce group differences as expected, the analysis needs to be re-run with a different number of clusters.

I used k-means clustering to cluster cases into three, four, five and six groups. As the resulting clusters did not reveal distinct differences in terms of one or two prominent criteria, I also used hierarchical clustering to form the clusters (explained below).

8.2.1 Hierarchical clustering

Like non-hierarchical clustering, *hierarchical* clustering forms clusters based on how similar or dissimilar the cases are. (However, unlike non-hierarchical clustering, hierarchical clustering does not start with an initial set of means and the number of clusters does not need to be predetermined.) Similarity or dissimilarity can be assessed in two ways – using distance measures or correlation coefficients. Distance measures, such as the Euclidean distance, calculate the distances between *all* pairs of cases, and then cluster groups accordingly. The shorter the distance, the more similar the cases; and in contrast, the further apart they are the more dissimilar the cases.[161] Correlation coefficients, on the other hand, measures the degree of similarity between cases – the higher the correlation coefficient, the more similar the cases.

If the range of criteria weights within cases is considered to be important, Pearson's correlation coefficient is the appropriate measure, whereas if the difference between cases is the main focus, a distance measure is more appropriate.

As I wanted to cluster the data based on the range of criteria weights *across* cases, I used the squared Euclidean distance measure which is the most commonly recognised distance measure and is particularly suitable for variables with low correlation and equal variances (Hair et al. 2010).[162]

The Euclidean distance uses the Pythagorean formula to measure the distance between two cases:

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

where x and y represent a pair of cases and i represents each respondent

[161] The distances between pairs of cases are commonly displayed in a symmetrical 'proximity matrix'.

[162] There are a variety of different distance measures including Euclidean distance, squared Euclidean distance, city-block (Manhattan) distance, Chebychev distance and Mahalanobis distance (D^2). Which distance measure is used depends on the data and the variables. For example, the Mahalanobis distance is the preferred option if standardisation of the data is required before cluster analysis can be undertaken, such as when the data are both categorical and numerical or if the range or scale of data is large (Hair et al 2010).

The squared Euclidean distance is the same as the Euclidean distance without taking the square root:

$$\sum_{i=1}^{n}(x_i - y_i)^2$$

I also used the Euclidean distance measure and Pearson's correlation coefficient in combination with different clustering methods in order to compare the results between the different methods (this is discussed further in the next section).

8.2.2 Formation of clusters

Clusters were formed using an agglomerative method. An agglomerative method starts with each case as an individual cluster. At each successive step, clusters are merged based on similarity, to form larger clusters. The process continues until there is one large cluster. Once a cluster is formed it cannot be split.[163]

There are a variety of clustering agglomerative algorithms used to combine clusters at each successive step, including 'single linkage', 'complete linkage', 'other linkage' (also known as the Centroid method), 'average linkage between groups', 'average linkage within groups' and 'Ward's method'. Similar to selecting a distance measure, selecting a linkage algorithm depends on the characteristics of the data. For example, using the 'single linkage' algorithm, based on minimum distances, often results in one large cluster with smaller clusters containing only a few cases whereas the 'complete linkage' algorithm, based on maximum distances, is strongly affected by outliers (Mooi & Sarstedt 2011). The two most commonly used clustering algorithms are 'average linkage' and 'Ward's method (Mooi & Sarstedt 2011).[164]

[163] The agglomerative method uses a 'bottom-up' approach whereas a divisive method uses a 'top-down' approach. A divisive method begins with one large cluster that includes all the cases. The large cluster is then split into smaller clusters that are the most similar. The process continues until all the cases are individual clusters. The agglomerative method is the most commonly used clustering method (Mooi & Sarstedt 2011).

[164] Average linkage *between* groups measures the average of all the possible distances between cases in each cluster whereas average linkage *within* groups combines clusters so that the average distance between all cases in the new cluster is as small as possible. Average linkage algorithms tend to produce clusters of similar sizes with low within-cluster variance (Mooi & Sarstedt 2011). Ward's method is based on minimising the sum of squares of any two clusters that can be formed at each step. Clusters are merged in such as way that variability within a cluster is minimised (Hair et al 2010).

Because 'average linkage' algorithms tend to be susceptible to outliers, I used the Ward's method in combination with a squared Euclidean distance measure to form clusters.[165] In order to compare the results using different methods, several other clustering algorithms were also used, in combination with two other similarity/dissimilarity measures: average linkage between groups and average linkage within groups (with a Pearson's correlation measure)[166] and average linkage within groups (with a Euclidean distance measure).

Deciding on the final number of clusters relies heavily on researcher judgement and is largely subjective (Norusis 2009). However, there are a number of clustering tools that can help in determining the appropriate number of clusters. These are discussed in the next section.

8.2.3 Determining the number of clusters

According to Norusis (2009) there is no set formula for determining the appropriate number of clusters to represent the data. Instead, the characteristics of each cluster need to be examined at each successive step to see if there is an "interpretable solution or a solution that has a reasonable number of fairly homogenous clusters" (Norusis 2009, p 361).

Clustering tools, including a dendrogram, an agglomeration schedule and a scree (or 'elbow') plot, can assist in determining the appropriate number of clusters (Norusis 2009).

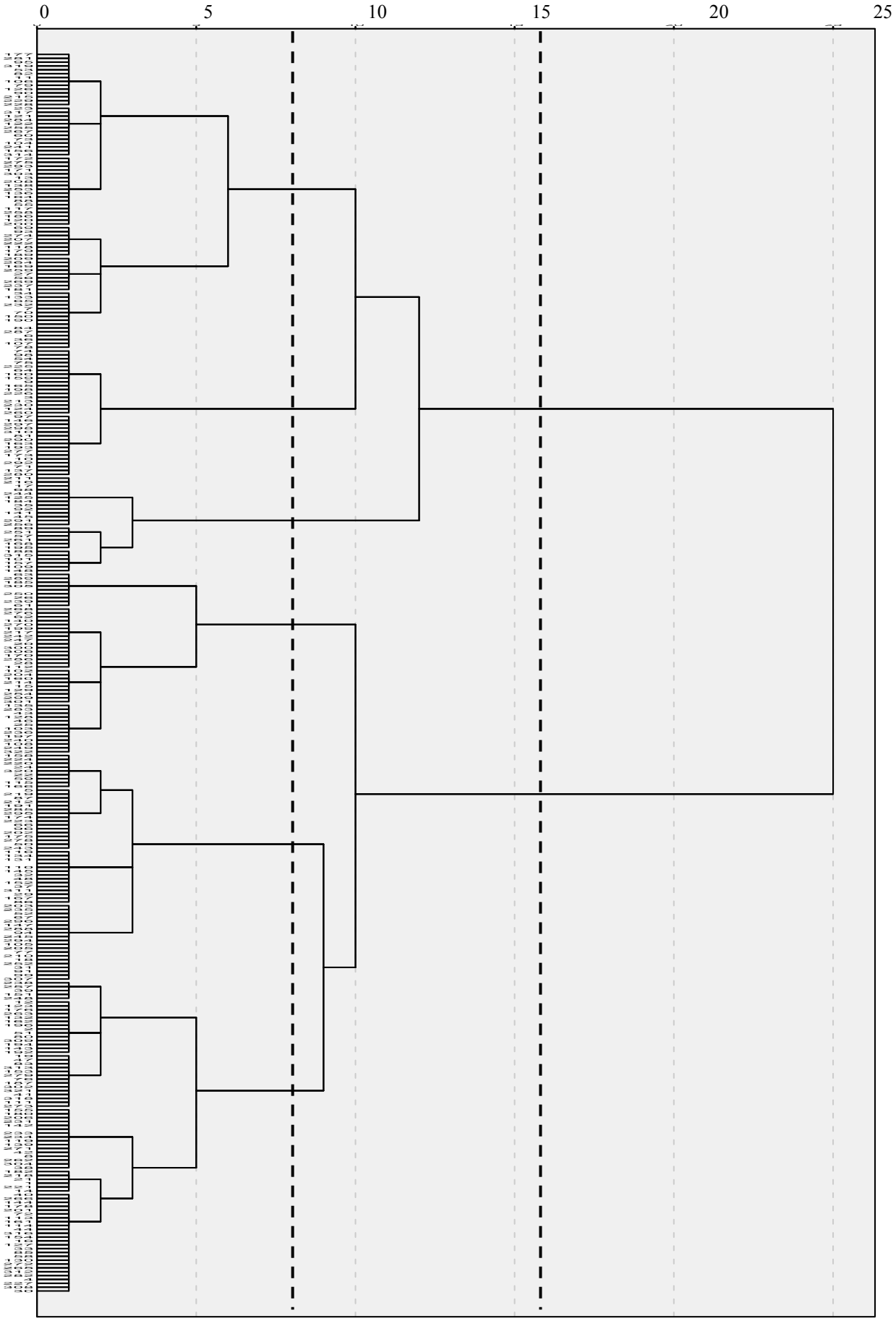
A dendrogram graphically illustrates when clusters are combined. Distances are rescaled to fall into the range of 1 to 25 with the ratio of the rescaled distances being the same as the ratio of the actual distances (Norusis 2009).

The dendrogram in Figure 8.1 illustrates how the respondents are clustered using Ward's method of clustering with a squared Euclidean distance measure.[167] The *solid* vertical lines indicate the distances at which the clusters are formed. Large distances between sequential vertical lines give some indication as to the appropriate number of clusters. The *dashed* vertical lines drawn on the dendrogram are at points where there is a comparatively large difference between the solid vertical lines. The number of times the dashed line cuts through the 'branches' provides an approximation of the suitable number of clusters. As can be seen in

[165] The squared Euclidean distance measure is recommended for Ward's method of clustering (Hair et al 2010).
[166] Correlation coefficients are recommended for clustering methods that are susceptible to outliers such as average linkage (Mooi & Sarstedt 2011).
[167] The dendrogram and agglomeration schedule (which will be explained shortly) were produced using SPSS (Statistical Package for Social Sciences) also known as PASW (Predictive Analytics Software).

Figure 8.1, dashed lines are drawn at distances of 7.5 and 15 where there are fairly large distances between the vertical lines. The dashed line at 7.5 crosses six branches indicating that six is an appropriate number of clusters. However, the dashed line at 15 crosses two branches suggesting that only two clusters are also suitable for the data.

Figure 8.1: Dendrogram



Another way to decide on the suitable number of clusters is to plot the coefficients from the agglomeration schedule against the number of clusters.

An agglomeration schedule, an example of which appears in Table 8.1 (showing the last 17 stages), illustrates the step-by-step clustering process of how the clusters are formed. The coefficients in the agglomeration schedule represent the value of the distance/similarity statistic being used. Cluster formation should stop when the increase (for distance measures) or decrease (for similarity measures) is large.

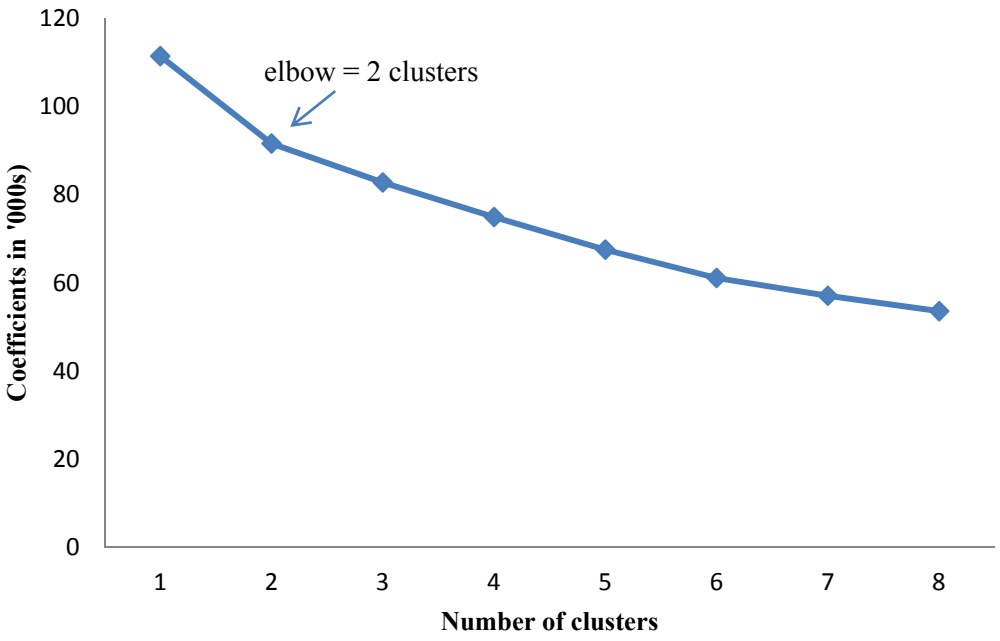
Table 8.1: Agglomeration schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
305	2	19	36319.804	278	281	308
306	6	69	37692.191	298	271	316
307	15	20	39151.035	299	285	315
308	2	39	40655.295	305	228	314
309	11	13	42230.223	276	303	316
310	18	29	43907.797	293	290	312
311	17	57	45921.210	291	301	320
312	5	18	47991.381	300	310	317
313	1	8	50189.055	302	277	314
314	1	2	53586.318	313	308	317
315	15	26	57084.993	307	288	319
316	6	11	61123.313	306	309	318
317	1	5	67551.062	314	312	319
318	3	6	74936.557	304	316	320
319	1	15	82789.612	317	315	321
320	3	17	91612.723	318	311	321
321	1	3	111433.672	319	320	0

The coefficients in an agglomeration schedule can be used to draw a scree plot. An ‘elbow’ in the scree plot indicates a large difference between clusters (Mooi & Sarstedt 2011). In Figure 8.2 the *last* 10 coefficients from the agglomeration schedule (using Ward’s method of clustering and the squared Euclidean distance) are plotted against the number of clusters. There is an ‘elbow’ at two indicating that two clusters are appropriate. Sometimes there is

more than one elbow in the graph which can make it difficult to identify the appropriate number of clusters.

Figure 8.2: Scree or 'elbow' plot



To ensure that the criteria weights are clustered as appropriately as possible, I used several different clustering methods and similarity/dissimilarity measures together with the clustering tools discussed above. A dendrogram, an agglomeration schedule and a scree plot were computed for each method and compared. Although the process is fairly subjective, between two and seven clusters appears to be an appropriate starting point.

There is a trade-off in deciding how many clusters to analyse. A large number of clusters will produce relatively homogenous-type clusters consisting of a small number of respondents whereas a small number of clusters will consist of a larger number of respondents but cluster membership may be fairly heterogeneous which limits analysis. The variation in respondents' preferences within clusters is therefore being traded-off with the variation of respondents' preferences between clusters.

The dendrogram and the scree plot (above) indicate that two clusters are appropriate. However, two clusters limit the analysis as the respondents are fairly heterogeneous in terms of their demographic characteristics. Therefore I clustered the data into two, three, four, five and six clusters for each method and explored the clusters further. For each cluster method, I

examined the individual clusters to determine whether there were any prominent groupings; for example, whether one cluster had a much higher mean weight for a particular criterion compared to the other clusters.

The different methods of clustering resulted in fairly similar clusters indicating that the clustering method and distance measure used in the analysis did not significantly affect the results. The Ward's method of clustering with a squared Euclidean measure gave the best results in terms of producing six reasonably sized clusters with each cluster showing a strong preference for one of the criteria. However, the Ward's method has a tendency to produce clusters of roughly the same size by splitting larger 'natural' clusters (Wilks 1995). Therefore the respondents are clustered twice; first into two clusters and then into six clusters and the membership of the both sets of clusters explored.

8.2.4 Membership of clusters

Using Ward's method of clustering with a squared Euclidean measure, the data were clustered into two clusters with 187 respondents in one cluster and 135 respondents in the other; and then into six clusters (ranging in size from 25 to 81 respondents). The mean criteria weights for each set of clusters are shown in Tables 8.2 and 8.3. A full demographic breakdown of the clusters is in Appendices 8.1 and 8.2.

Table 8.2: Mean criteria weights for two clusters

Criteria	Cluster 1 (n = 187)	Cluster 2 (n = 135)
Only available treatment	11.9%	8.7%
Societal benefit	11.1%	13.5%
Lifestyle	10.6%	15.8%
Age	11.8%	17.5%
Need	33.8%	20.8%
Individual benefit	20.8%	23.7%

As can be seen in Table 8.2 the mean criteria weights are quite different between the two clusters. Compared to Cluster 1, Cluster 2 has higher mean criteria weights for four of the criteria including 'lifestyle' (+5.2%) and 'age' (+5.7%) and a much lower mean criteria weight for 'need' (−13%).

Table 8.3: Mean criteria weights for six clusters

	Cluster 1 (*n* = 81)	Cluster 2 (*n* = 33)	Cluster 3 (*n* = 59)	Cluster 4 (*n* = 77)	Cluster 5 (*n* = 47)	Cluster 6 (*n* = 25)
Only available treatment	7.0%	8.7%	**18.5%**	8.5%	12.2%	9.3%
Age	11.1%	16.9%	11.3%	13.7%	13.5%	**29.6%**
Societal benefit	14.4%	**19.7%**	10.7%	13.3%	5.9%	6.0%
Need	33.5%	19.6%	27.7%	19.7%	**42.1%**	26.1%
Individual benefit	23.8%	**29.1%**	20.8%	23.5%	15.6%	17.4%
Lifestyle	10.2%	6.0%	11.0%	**21.4%**	10.7%	11.7%

Note: the numbers in bold indicate criteria weights which are much higher than in other clusters

When respondents are clustered into six clusters, the mean criteria weights in Cluster 1, the largest cluster, closely resemble the overall mean weights with 'need' and 'individual benefit' the most important criteria and 'only available treatment' the least important. Cluster 2, the second smallest cluster, considers 'individual benefit' to be the most important criterion. The weights for 'individual benefit' and 'societal benefit' in Cluster 2 are higher than for any other cluster, and combined, make up almost 50% of the total criteria weights within the cluster.

In Cluster 3 the mean weight for 'only available treatment' is much higher than for any other cluster. Relative to all other clusters, respondents in Cluster 4, the second largest cluster, place relatively more importance on 'lifestyle'. Cluster 4 is the only cluster that has a higher mean weight for 'lifestyle' than 'need'.

Cluster 5 has the highest mean weight for 'need' and the lowest mean weights for 'societal benefit' and 'individual benefit' compared to the other five clusters. In Cluster 5 the mean weight for 'need' is seven times higher than the mean weight for 'societal benefit'.

Cluster 6, the smallest cluster with 25 respondents, has a mean weight for 'age' which is much higher than all other clusters, and like Cluster 5, has low mean weights for 'societal benefit' and 'individual benefit'. In Cluster 6 the mean weight for 'age' is almost six times higher than the mean weight for 'societal benefit'.

The next step is to see whether the respondents in the clusters share specific demographic characteristics.

8.3 Exploring cluster membership

To find out whether there are demographic characteristics common to specific clusters, such as age or income for example, chi-square (χ^2) tests are performed. In the following sections, the chi-square test is explained and the results of the analyses are presented.

8.3.1 The chi-square test of independence

The chi-square test of independence (also known as the Pearson's chi-square test) tests the association between two categorical variables. The chi-square statistic (χ^2) is calculated by summing the difference between the expected and observed frequencies for each level of a categorical variable and dividing by the expected frequencies (Norusis 2009).

$$\chi^2 = \Sigma\frac{(\text{observed}_{ij} - \text{expected}_{ij})^2}{\text{expected}_{ij}}$$

where i = rows and j = columns in a contingency table; expected $_{ij}$ = (row total$_i$ × column total$_j$)/n

For example, 59.6% of the random sample respondents are female and 40.4% are male. Therefore the expected frequency of females in each cluster is 59.6%. However, the observed frequency may be different. The chi-square test is used to test whether there is an association between 'gender' and each cluster.

Specifically, the χ^2 statistic is compared to an estimate of the probability of obtaining χ^2 if there is no association between the variables. If the statistic is less than 0.05 the observed frequencies are judged to be different from the expected frequencies and therefore an association between the two variables exists (Norusis 2009).

Two assumptions need to be met for the chi-square test to be valid. First, the observations or scores need to be independent of each other. Second, because the significance tests of the chi-square distribution may not be accurate with smaller samples,[168] all expected cell frequencies in a 2×2 table[169] should be greater than five, or for larger tables, 80% of the cells should have an expected cell count greater than five (Field 2009). The sampling distribution will then tend towards a perfect chi-square distribution. Fisher's exact test can also be used to compute the probability of the chi-square statistic when sample sizes are small or the expected frequencies

[168] Low expected counts can lead to an inflated χ^2 statistic.
[169] A 2×2 table has two categorical variables (e.g. 'gender' and 'cluster'), both of which have two categories (i.e. 'male' or 'female', and 'in the cluster' or 'not in the cluster').

are too low (Field 2009). An example of a 2×2 contingency table and the associated chi-square tests are shown in Tables 8.4 and 8.5.

Table 8.4: 2×2 contingency table for the category 'gender'

			Gender		
			Male	Female	Total
Cluster 1	Not Selected	Count	96	145	241
		Expected Count	97.3	143.7	241.0
		% within Cluster 1	39.8%	60.2%	100.0%
		% within Gender	73.8%	75.5%	74.8%
		% of Total	29.8%	45.0%	74.8%
		Std. Residual	−0.1	0.1	
	Selected	Count	34	47	81
		Expected Count	32.7	48.3	81.0
		% within Cluster 1	42.0%	58.0%	100.0%
		% within Gender	26.2%	24.5%	25.2%
		% of Total	10.6%	14.6%	25.2%
		Std. Residual	0.2	−0.2	
Total		Count	130	192	322
		Expected Count	130.0	192.0	322.0
		% within Cluster 1	40.4%	59.6%	100.0%
		% within Gender	100.0%	100.0%	100.0%
		% of Total	40.4%	59.6%	100.0%

In the last section of Table 8.4, it can be seen that 130 respondents (40.4%) from the random sample are male and 192 (59.6%) are female. Therefore the expected number of males in Cluster 1 should equate to 40.4% of the total number of respondents in Cluster 1 and the expected number of females should equate to 59.6% of the total number of respondents in Cluster 1. In the second section of Table 8.4 it can be seen that the expected number of males in Cluster 1 is 32.7 (40.4% of 81) and the expected number of females is 48.3 (59.6% of 81).[170] However, the actual (or observed) number of males in Cluster 1 is 34 and the number of females is 47. (The 'Std. Residual' in Table 8.4 will be explained shortly.)

[170] Statistically, the expected numbers are 32.7 and 48.3 but as gender is not divisible the expected number of males and females is *approximately* 33 and 48 respectively!

Table 8.5: Table of chi-Square test results for 'gender'

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	0.115[a]	1	0.734		
Continuity Correction[b]	0.044	1	0.835		
Likelihood Ratio	0.115	1	0.734		
Fisher's Exact Test				0.794	0.416
N of Valid Cases	322				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 32.70.
b. Computed only for a 2x2 table

As explained previously, to determine whether there is an association between 'gender' and Cluster 1, the chi-square statistic is compared to the probability of obtaining χ^2 if there were no association between the two groups. As can be seen in Table 8.5, the chi-square statistic (χ^2 = 0.734) is greater than 0.05. Therefore there is no association between 'gender' and Cluster 1.

With 2×2 tables, Pearson's chi-square is inclined to underestimate the true significance values (particularly when cells have an expected count of less than five), increasing the risk of a Type I error (Field 2009). To adjust for this, the Yate's continuity correction can be used.

$$\chi^2 = \Sigma \frac{(\text{observed}_{ij} - \text{expected}_{ij} - 0.5)^2}{\text{expected}_{ij}}$$

However, Field (2007) warns that there is evidence to suggest that this solution can overcorrect and produce a chi-square value that is too low.

If the chi-square test is significant, the strength of the association between the variables is determined by calculating the effect size. Two commonly used statistics are the phi coefficient for 2×2 tables and Cramer's V for larger tables.[171] The absolute value of the phi coefficient and Cramer's V lies between 0 and 1, with 0 indicating no correlation and 1 indicating perfect correlation. The size of the effect can be interpreted using Cohen's (1988) criteria: 0.10 for small effect, 0.30 for medium effect and 0.50 for large effect.[172]

[171] For 2×2 tables, the phi coefficient and Cramer's V will be the same.
[172] These criteria are used for interpreting the phi coefficient. For Cramer's V (associated with larger tables) the effect size interpretation differs depending on the number of categories. However, because each cluster is being

A significant chi-square statistic does not identify which of the cross-tabulated groups is under or over-represented with respect to the expected frequencies. The contingency table needs to be examined to determine which cell or cells have produced the statistically significant difference. However, interpreting significance by examining the percentages in the contingency table may not be accurate (Field 2009). Examining the standardised residuals is a more reliable indicator. A standardised residual is the difference between the observed frequency and the expected frequency converted to a *z*-score. A standardised residual is significant if it is greater than ±1.96 at $p < 0.05$, ±2.58 at $p < 0.01$ and ±3.29 at $p < 0.001$ (Field 2009). Positive values mean that the observed frequencies are greater than the expected frequency, whereas negative values mean that the observed frequencies are less than the expected frequencies.

For example, when tabulating Cluster 3 and 'age group' the chi-square statistic is significant, indicating a relationship between Cluster 3 and 'age group'. As can be seen in Table 8.6 there are two statistically significant standardised residuals (in bold). These statistically significant residuals indicate which groups have produced the statistically significant chi-square statistic. In this example, respondents in Cluster 3 aged 25-34 years are under-represented (−2.0) and respondents aged 55-64 (2.0) are over-represented.

considered separately, the row size will always be two (respondents in a cluster/respondents not in a cluster) which means that Cohen's criteria can be used to estimate the effect size.

Table 8.6: 2×6 contingency table for Cluster 3 and 'age'

Cluster 3		Age group						
		18-24	25-34	35-44	45-54	55-64	65 yrs +	Total
Not Selected	Count	17	30	33	66	64	53	263
	Expected Count	18.8	25.3	34.3	60.4	71.9	52.3	263.0
	% within Cluster 3	6.5%	11.4%	12.5%	25.1%	24.3%	20.2%	100.0%
	% within Age group	73.9%	96.8%	78.6%	89.2%	72.7%	82.8%	81.7%
	% of Total	5.3%	9.3%	10.2%	20.5%	19.9%	16.5%	81.7%
	Std. Residual	−0.4	0.9	−0.2	0.7	−0.9	0.1	
Selected	Count	6	1	9	8	24	11	59
	Expected Count	4.2	5.7	7.7	13.6	16.1	11.7	59.0
	% within Cluster 3	10.2%	1.7%	15.3%	13.6%	40.7%	18.6%	100.0%
	% within Age group	26.1%	3.2%	21.4%	10.8%	27.3%	17.2%	18.3%
	% of Total	1.9%	.3%	2.8%	2.5%	7.5%	3.4%	18.3%
	Std. Residual	0.9	**−2.0**	0.5	−1.5	**2.0**	−0.2	
Total	Count	23	31	42	74	88	64	322
	Expected Count	23.0	31.0	42.0	74.0	88.0	64.0	322.0
	% within Cluster 3	7.1%	9.6%	13.0%	23.0%	27.3%	19.9%	100.0%
	% within Age group	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	% of Total	7.1%	9.6%	13.0%	23.0%	27.3%	19.9%	100.0%

Note: the bold numbers indicate statistically significant standardised residuals

8.3.2 Results of the chi-square tests of independence

In Tables 8.7 and 8.8 the results of statistically significant chi-square tests (at the 5% level) are presented.[173] The demographic characteristics of the respondents were cross-tabulated with each cluster to see if the respondents in a cluster had characteristics in common. For many of the demographic characteristics that consisted of more than two groups (for example, six income groups), expected cell sizes were too small to continue with the analysis. Where practicable, the groups were amalgamated to form a smaller number of groups to ensure 80% of the cell sizes were above five. For example, 'region' which originally had 13 groups was collapsed into two groups: North Island and South Island. The data were also spilt by gender and separate analyses conducted.

[173] The full sets of results are in Appendices 8.3 and 8.4.

Table 8.7: Results of chi-square testing (two clusters)[174]

	Independent variables	Significance	Effect	Significant group (Std. Residual > 1.96)
Cluster 1 (*n*=187) (58 %) **compared with** **Cluster 2 (*n*=135) (42 %)**	Children/no children	χ^2 (1, *n*=187) = 6.270, *p*=0.012	small	none
	Regions, *male only* (South Island/North Island)	χ^2 (1, *n*=187) = 8.229, *p*=0.004	small	none
	Serious illness (yes/no)	χ^2 (1, *n*=187) = 6.270, *p*=0.012	small	none
	Serious illness, *female only* (yes/no)	χ^2 (1, *n*=187) = 4.801, *p*=0.028	small	none
	Work status (not working/working/retired)	χ^2 (2, *n*=187) = 12.457, *p*=0.002	small-medium	Not working (over-represented in Cluster 1)
	Work status, *female only* (not working/working/retired)	χ^2 (2, *n*=187) = 13.663, *p*=0.001	medium	Not working (over-represented in Cluster 1)

Notes: Pearson chi-square (χ^2) reported (only reported if Yate's continuity correlation significant); Effect sizes using Cohen's (1988) criteria: 0.1=small; 0.3=medium; 0.5=large.

When there are only two clusters, if an association is found between Cluster 1 and one of the demographic characteristics, it means there is also an association between Cluster 2 and the same demographic characteristic. For example, if the number of respondents who do not work are *over*-represented in Cluster 1, it means that the number of respondents who do not work are *under*-represented in Cluster 2.

As can be seen in Table 8.7, Pearson's chi-square testing indicates an association between each cluster and the following demographic categories: children/no children, serious illness, work status (not working/working/retired) and the region in which male respondents reside, with all associations showing a small effect.

As mentioned previously, compared to Cluster 2, Cluster 1 has lower mean criteria weights for four of the criteria ('individual benefit', 'societal benefit', 'age' and 'lifestyle) and a much higher mean criteria weight for 'need' (+13%). The work status of females (not working, working and retired) is also associated with the clusters, with a significantly higher number of non-working female respondents than expected in Cluster 1 (and conversely, a significantly lower number of non-working female respondents in Cluster 2), with a medium effect.

[174] The Yate's continuity correction significance level is used for 2×2 contingency tables. Demographic characteristics showing significance using the Pearson's chi-square statistic still show significance using the more conservative Yate's continuity correction. Cramer's V is used to determine the effect size and the standardised residual is used to determine which of the groups (of the demographic category) has the most influence on the significant chi-square test statistic.

The results of the chi-square testing for the six clusters are presented in Table 8.8 and discussed below.

Table 8.8: Results of chi-square testing (six clusters)

	Independent variables	Significance	Effect	Significant group (Std. Residual > 1.96)
Cluster 1 (*n*=81) (25%)	Health insurance, *male only* (yes/no)	χ^2 (1, *n*=34) = 5.178, *p*=0.023	small	none
	Health usage (seldom/occasionally/frequently)	χ^2 (2, *n*=80) = 6.153, *p*=0.048	small	none
	Serious illness (yes/no)	χ^2 (1, *n*=81) = 6.017, *p*=0.014	small	none
	Serious illness, *female only* (yes/no)	χ^2 (1, *n*=47) = 4.664, *p*=0.031	small	none
	Work status (not working/working/retired)	χ^2 (2, *n*=77) = 9.667, *p*=0.008	small	Retired group (under-represented)
	Work status, *female only* (not working/working/retired)	χ^2 (2, *n*=45) = 10.531, *p*=0.005	medium	Retired group (under-represented)
	Work status, *female only* (working/not working)	χ^2 (1, *n*=45) = 6.082, *p*=0.014	small	none
Cluster 2 (*n* = 33) (10%)	Children/no children, *female only*	χ^2 (1, *n*=17) = 5.587, *p*=0.018	small	none
Cluster 3 (*n*=59) (18%)	Age (six age groups)	χ^2 (5, *n*=59) = 13.474, *p*=0.019	medium	25-35 age group (under-represented) 55-64 age group (over-represented)
	Work status, *female only* (working/not working)	χ^2 (1, *n*=39) = 5.414, *p*=0.020	small	none
Cluster 4 (*n*=77) (24%)	Regions, *male only* (South Island/North Island)	χ^2 (1, *n*=32) = 4.893, *p*=0.027	small	none
	Work status (not working/working/retired)	χ^2 (2, *n*=75) = 9.027, *p*=0.011	small	Not working (under-represented)
	Work status, *female only* (not working/working/retired)	χ^2 (2, *n*=45) = 7.687, *p*=0.021	medium	Not working (under-represented)
Cluster 5 (*n*=47) (15%)	Qualifications (none/secondary school and post secondary school)	χ^2 (1, *n*=47) = 4.568, *p*=0.033	small	none
	Regions, *male only* (South Island/North Island)	χ^2 (1, *n*=22) = 6.581, *p*=0.010	medium	South Island (over-represented)
	Stated income (six income groups)	χ^2 (5, *n*=39) = 11.928, *p*=0.036	medium	None
Cluster 6 (*n*=25) (8%)	Qualifications (none/secondary school and post secondary school)	χ^2 (1, *n*=25) = 4.956, *p*=0.026	small	none
	Qualifications, *female*	χ^2 (1, *n*=15) = 7.028, *p*=0.005	medium	No qualifications/ secondary school (over-represented)

Notes: Pearson chi-square (χ^2) reported (only reported if Yate's continuity correlation significant); Effect sizes using Cohen's (1988) criteria: 0.1=small; 0.3=medium; 0.5=large.

In Cluster 1 the two criteria with the highest mean weights are 'need' and 'individual benefit'. Pearson's chi-square testing indicates an association between Cluster 1 with the following

demographic categories: health insurance (males only), frequency of using health services, experience of a serious illness and the work status of females (working/not working), all demonstrating a small effect. Work status, divided into three groups (not working, working and retired) is also associated with Cluster 1, with a significantly lower number of retired respondents than expected in this cluster. When work status (not working, working and retired) is analysed by gender, retired female respondents are significantly under-represented in Cluster 1, with a medium effect.

Cluster 2's combined criteria weights for 'societal benefit' and 'individual benefit' are much higher than the combined weights for these criteria in any other cluster. The only association between Cluster 2 and the demographic characteristics is female respondents who live with children, showing a small effect.

The highest mean criterion weight for 'only available treatment' is in Cluster 3. Females' work status is associated with Cluster 3, with a small effect. An association with age (when split into six groups) is also found. The number of respondents aged 25-35 years is significantly *less* than expected and the number of respondents aged 55-64 years is significantly *more* than expected, with a medium effect.

The mean weight for 'lifestyle' is considerably higher in Cluster 4 than in any other cluster. The region in which males live (North Island /South Island) and females' work status (working/not working) are found to be associated with Cluster 4, both showing a small effect. Work status overall (not working, working and retired) is also associated with Cluster 4, with respondents who do not work being significantly under-represented. When analysing work status by gender, females who do not work are also significantly under-represented, with a medium effect.

Compared to the other clusters, the mean weight for 'need' is much higher in Cluster 5. When the demographic category 'qualifications' is collapsed from four groups to two[175] an association is found between the two qualification groups and Cluster 5, with a small effect. Income (in six groups) is also associated with Cluster 5, showing a medium effect. The region in which males live is associated with Cluster 5, with more South Island males than expected, in the cluster.

[175] No qualifications/secondary school and post secondary school qualifications.

The smallest cluster, Cluster 6, has the highest mean criterion weight for 'age', almost twice as high as any other cluster. Qualifications are associated with this cluster with the number of females who are not qualified or have secondary school qualifications only, being over-represented in this cluster, showing a medium effect.

8.3.3 Summary of cluster analysis and chi-square testing

When the respondents are clustered into two clusters there are several associations between the clusters and the demographic characteristics of the respondents. However, the only demographic characteristic that has a statistically significant residual is 'work status for females'. Cluster 1, which has a much larger mean criterion weight for 'need' compared to Cluster 2, has a lower number of non-working female respondents in the Cluster than expected.

When the respondents are clustered into six clusters Cluster 1's criteria weights closely resemble the *overall* criteria weights for the random sample. Every other cluster has a relative preference for one of the criteria (or two criteria in the case of Cluster 2). When analysing the clusters, however, the main demographic characteristics such as age, ethnicity, income and qualifications, are not highly associated with the clusters. The characteristics that are associated with the clusters are health insurance, health usage, experience of a serious illness and work status.

Many of the clusters have only a small number of respondents, particularly when the cluster is analysed by gender alone. For example, Cluster 6 is associated with females who have either secondary school qualifications or no qualifications, but there are only 15 females in the cluster (i.e. 4.6% of the random sample). Similarly, there is an association between Cluster 5 and the region in which males live but there are only 22 males in the cluster (i.e. 6.8% of the random sample). In addition, the effect sizes for most of the associations are small, indicating that, although several demographic characteristics are associated with specific clusters, the association is minimal.[176]

[176] Of the 18 significant associations between the six clusters and the demographic characteristics of the respondents, only seven have a significant standardised residual. (A significant standardised residual indicates which of the sub-groups has the most effect on the cluster.)

Though there are some statistically significant associations between the clusters and the criteria, the effects are mainly small and the associations generally involve a small number of respondents.

However, given that there are differences in the mean criteria weights across clusters, what effect does this have on prioritising health treatments? This is discussed in the next section.

8.4 Ranking vignettes using cluster weights

When the data are split into six clusters, the mean weights for some criteria differ markedly across the clusters. In this section, the mean criteria weights for each cluster are used to rank the 14 health treatments from the focus groups, to explore what happens to overall treatment rankings when different criteria weights are used.

To start, the 14 health treatments are categorised according to the criteria and levels.[177] For example, in Figure 8.3 'dialysis for renal disease' is categorised: it is the only treatment available; renal disease predominantly affects patients over 65 years; dialysis provides a small benefit to others and so on.

Figure 8.3: Categorising 'dialysis for renal disease'

	CRITERIA					
ALTERNATIVE click to open	Treatment options for this patient	Age of patient	Benefit to others (*eg* family or society)	Patient's health before treatment	Benefit to patient (*ie* length and/or quality of life)	Illness or injury caused mainly by lifestyle choices
Dialysis for renal disease	this is the **ONLY** treatment available	65+ years	small	will die soon without treatment	large	yes

Each health treatment is then 'scored' using the criteria weights.[178] For example, in Figure 8.4, Cluster 1's criteria weights are used to score 'dialysis for renal disease': as there are no other treatment options for this patient, the treatment scores '7.0%' on the first criterion; the average age of a patient is over 65 years therefore the treatment scores '0%' on the second

[177] I categorised the health treatments using all available information including advice from health experts. However, the categorising and scoring of health treatments is for illustrative purposes only.
[178] This is done automatically by 1000Minds software (Hansen & Ombler 2009).

criterion and so on.[179] The scores on each criterion are added together to get an overall score for the health treatment.

Figure 8.4: Scoring 'dialysis for renal disease' using Cluster 1's criteria weights

Treatment options for this patient		Score
this is the best treatment (there are less effective alternatives)	0.0 %	7.0
this is the **ONLY** treatment available	7.0 %	
Age of patient		
65+ years	0.0 %	0
15-64 years	5.9 %	
0-14 years	11.1%	
Benefit to others (*eg* family or society)		
small	0.0 %	0
large	14.4 %	
Patient's health before treatment		
relatively good (though treatment is still beneficial)	0.0 %	33.5
fair (neither good nor bad)	8.4%	
poor (but not immediately life threatening)	17.3%	
will die soon without treatment	33.5 %	
Benefit to patient (*ie* length and/or quality of life)		
small	0.0 %	23.8
medium	13.7 %	
large	23.8 %	
Illness or injury caused mainly by lifestyle choices		
yes	0.0 %	0
no	10.2 %	
	Total Score:	64.3

Once all the treatments are categorised and scored, the treatments can be ranked according to the total scores. Figure 8.5 displays the 14 treatments ranked using Cluster 1's criteria weights.[180]

[179] As discussed in Chapter 6, the total score is sensitive to changes in the criteria and/or levels. For example, if the age of the patient was 15-64, then the score on the 'age' criterion would be 5.9% and not 0%.

Figure 8.5: Treatment ranks using Cluster 1's criteria weights

	CRITERIA							
TECHNOLOGY click to open	Treatment options for this patient	Age of patient	Benefit to others (*eg* family or society)	Patient's health before treatment	Benefit to patient (*ie* length and/or quality of life)	Illness or injury caused mainly by lifestyle choices	RANK	TOTAL SCORE
Dialysis for renal disease	this is the ONLY treatment available	65+ years	small	will die soon without treatment	large	yes	1st	64.3 %
Hip replacements	this is the ONLY treatment available	15-64 years	small	poor (but not immediately life threatening)	large	no	2nd	64.3 %
Imatinib mesylate for chronic myeloid leukaemia	this is the best treatment (there are less effective alternatives)	15-64 years	small	will die soon without treatment	medium	no	3rd	63.3 %
Antiretroviral drugs for HIV	this is the ONLY treatment available	15-64 years	small	poor (but not immediately life threatening)	large	yes	4th	54.0 %
Abatacept for rheumatoid arthritis	this is the best treatment (there are less effective alternatives)	15-64 years	small	poor (but not immediately life threatening)	medium	no	5th	47.1 %
Service for postnatal depression	this is the best treatment (there are less effective alternatives)	15-64 years	large	fair (neither good nor bad)	small	no	6th	38.9 %
Methadone	this is the best treatment (there are less effective alternatives)	15-64 years	large	poor (but not immediately life threatening)	small	yes	7th	37.6 %
Growth hormone treatment for Prader-Willi Syndrome	this is the ONLY treatment available	0-14 years	small	fair (neither good nor bad)	small	no	8th	36.7 %
Hand sanitiser use in primary schools	this is the best treatment (there are less effective alternatives)	0-14 years	large	relatively good (though treatment is still beneficial)	small	no	9th	35.7 %
PET Scan	this is the best treatment (there are less effective alternatives)	15-64 years	small	fair (neither good nor bad)	small	no	10th	24.5 %
IVF Treatment	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	16.2 %
Oral drugs for erectile dysfunction	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	16.2 %
Vaccine for cervical cancer	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	16.2 %
Statins for cardiovascular disease	this is the best treatment (there are less effective alternatives)	15-64 years	small	fair (neither good nor bad)	small	yes	14th	14.3 %

The same process is used to rank the 14 treatments using each of the clusters' criteria weights. In Table 8.9 the 14 treatments are ranked using the mean weights from each cluster, the random sample mean criteria weights, 'evenly assigned' weights and 'arbitrarily assigned' weights.

[180] In Figure 8.5 'dialysis for renal disease' is ranked first even though it has the same score as 'hip replacements' which is second (64.3%). This is due to rounding errors. 'Dialysis for renal disease' has a slightly higher score (64.279) than 'hip replacements' (64.251).

Table 8.9: Rankings of health treatments by the six clusters, random sample and arbitrarily assigned weights

Health treatments	Clusters						Random sample	Criteria weights evenly assigned*	Criteria weights arbitrarily assigned**
	1	2	3	4	5	6			
Hip replacements	2	1	1	1	2	1	1	1	1
Dialysis for end-stage renal disease	1	2	2	3	1	5	2	5	9
Imatinib mesylate for chronic myeloid leukaemia	3	4	4	2	3	2	3	4	5
Antiretroviral drugs for HIV	4	3	3	6	4	4	4	5	6
Abatacept for last-line treatment of rheumatoid arthritis	5	6	6	4	5	6	5	8	6
Growth hormone for Prader-Willi Syndrome	8	9	5	7	6	3	6	2	3
Hand sanitiser use in primary schools	9	5	7	5	9	7	7	3	1
Service for postnatal depression	6	7	8	8	8	8	8	7	4
Methadone for opioid addiction	7	8	9	10	7	9	9	9	8
Positron emission tomography (PET Scan)	10	10	10	9	10	10	10	10	10
Vaccine for preventing cervical cancer	11	11	11	11	11	11	11	11	11
IVF treatment	11	11	11	11	11	11	11	11	11
Oral drugs for erectile dysfunction	11	11	11	11	11	11	11	11	11
Statins for patients at high risk of cardiovascular disease	14	14	14	14	14	14	14	14	14

*Criteria weights evenly assigned – each criterion has been assigned the average criterion weight, ie 100/6 = 16.6 or 16.7.

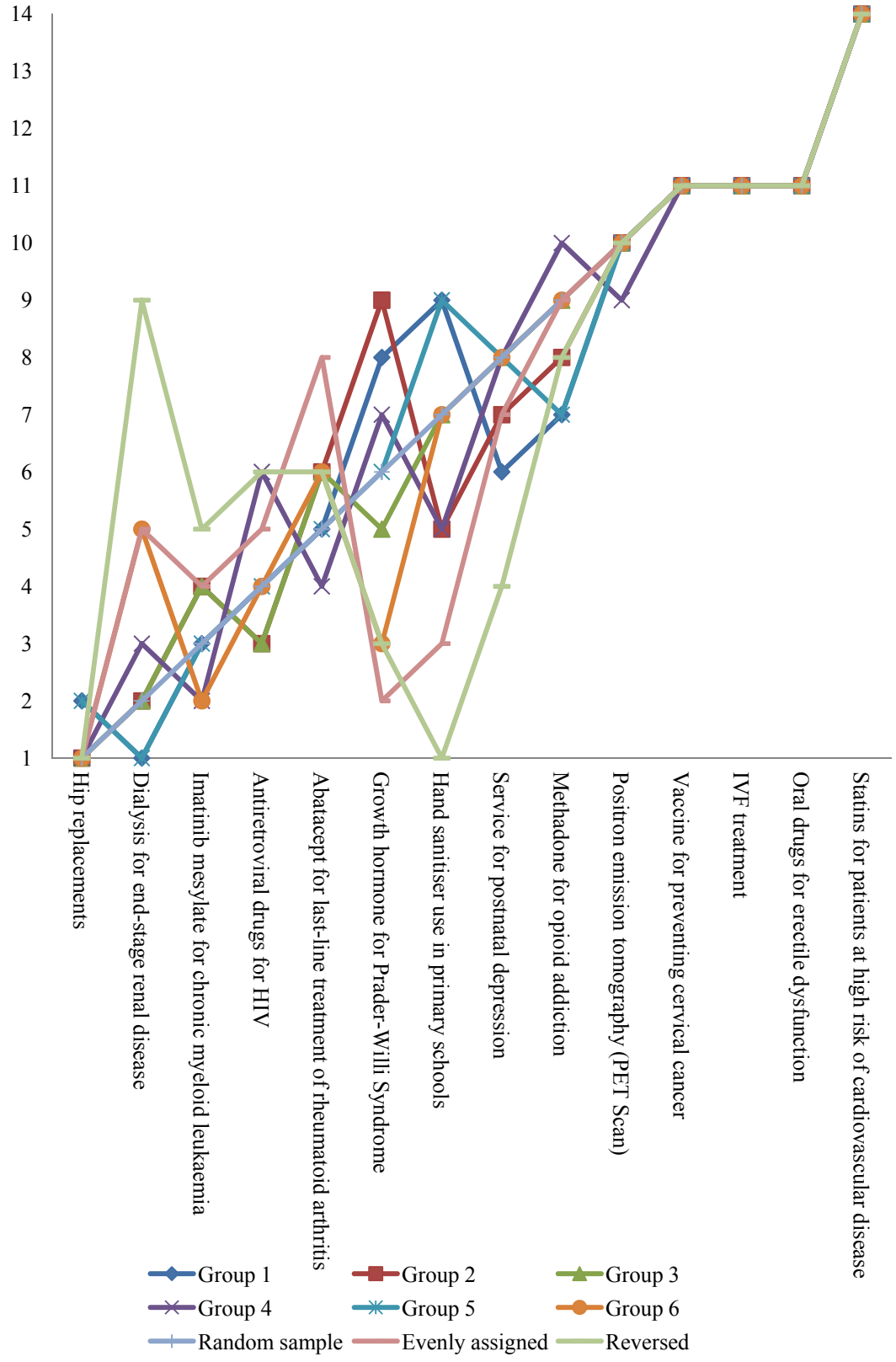
**Criteria weights arbitrarily assigned – the first number in parentheses is the 'arbitrarily assigned' criteria weight and the second number is the random sample mean criteria weight (to compare) – 'need' (12, 28.4), 'individual benefit' (10, 22), 'age' (23, 14.2), 'lifestyle' (20, 12.8), 'societal benefit' (20, 12.1) and 'no alternative treatment' (15, 10.5).

The last two sets of weights are used to illustrate the extent to which rankings change when criteria weights vary across groups. Before discussing this further, I will explain what is meant by 'evenly assigned' and 'arbitrarily assigned' weights. In the second last column in Table 8.9 'evenly assigned' criteria weights have been used to rank the health treatments. This means that the highest level of each criterion is assigned a weight of either 16.6 or 16.7, which is the total amount of points divided by the number of criteria (i.e. 100/6) and the lower levels of each criterion are assigned proportionately lower weights which are evenly distributed across the levels. In the last column of Table 8.9 'arbitrarily assigned' criteria weights are used to rank the health treatments. This means that the random sample criteria

weights have been arbitrarily ‘re-assigned’ with the higher-weighted criteria from the random sample receiving lower weights and the lower-weighted criteria from the random sample receiving higher weights. For example, the highest random sample mean criterion weight is for ‘need’ (i.e. 28.4%). The ‘arbitrarily assigned’ weight for ‘need’ is 12% whereas the lowest random sample mean criterion weight is for ‘no alternative treatment’ (i.e. 10.5%) and the ‘arbitrarily assigned’ weight for ‘no alternative treatment’ is 15%. Weights for the lower levels of the criteria are then proportionately re-assigned in line with the highest level weights.

The information from Table 8.9 is displayed graphically in Figure 8.6.

Figure 8.6: Graph of rankings of health treatments by the six clusters, random sample and arbitrarily assigned weights



Rankings of the treatments are affected by two factors – the categorisation of the treatments and the criteria weights. To illustrate these points, in addition to using the mean weights from the clusters and the overall random sample to rank the treatments, the two arbitrarily assigned sets of criteria weights are used to rank the treatments (i.e. the last two columns of Table 8.9).

The categorisation of health treatments influences the final rankings regardless of the criteria weights used. When a treatment is categorised as having the highest level for *all* of the criteria (for example, 'large' on the criterion 'benefit to patient', '0-14' on the criterion 'age' and so on), then that treatment will be ranked first regardless of the associated criteria weights. Similarly, when a treatment is categorised at the highest level for *most* of the criteria then that treatment will score highly regardless of the criteria weights. For example, 'hip replacements' is categorised at the highest level for four of the criteria, and at the second highest level for two of the criteria. As can be seen in Table 8.9, apart from Cluster 1, 'hip replacements' is ranked first. In contrast, 'statins' is categorised at the lowest level for four of the criteria, and at the second lowest level for two of the criteria. 'Statins' is ranked last regardless of the criteria weights used. As 'vaccine for cervical cancer', 'IVF' and 'oral drugs for erectile dysfunction' are categorised exactly the same, they are equally ranked (11^{th}=) regardless of the criteria weights used.

However, when treatments are categorised as having a mixed range of levels, criteria weights affect the rankings. The criteria weights for the six clusters are displayed in Table 8.10.

Table 8.10: Mean criteria weights for each cluster

	Cluster 1 (n = 81)	Cluster 2 (n = 33)	Cluster 3 (n = 59)	Cluster 4 (n = 77)	Cluster 5 (n = 47)	Cluster 6 (n = 25)
Only available treatment	7.0%	8.7%	**18.5%**	8.5%	12.2%	9.3%
Age	11.1%	16.9%	11.3%	13.7%	13.5%	**29.6%**
Societal benefit	14.4%	**19.7%**	10.7%	13.3%	5.9%	6.0%
Need	33.5%	19.6%	27.7%	19.7%	**42.1%**	26.1%
Individual benefit	23.8%	**29.1%**	20.8%	23.5%	15.6%	17.4%
Lifestyle	10.2%	6.0%	11.0%	**21.4%**	10.7%	11.7%

The bolded numbers indicate a stronger preference for that particular criterion compared to the other clusters.

As most clusters consider 'need' and 'individual benefit' to be the most important criteria, treatments associated with a high health need and/or a large individual benefit, such as

'dialysis for renal disease', 'hip replacements', 'imatinib mesylate for chronic myeloid leukaemia' and 'antiretroviral drugs for HIV' are ranked in the top four by most of the clusters.

When Cluster 4's criteria weights are used to rank the treatments 'antiretroviral drugs for HIV' and 'methadone' are ranked lower compared to the other clusters. This is because the mean criterion weight for 'lifestyle' in Cluster 4 is almost double the weight of any other cluster – treatments that are *not* associated with poor lifestyle choice score 21.4% on 'lifestyle' alone. As 'antiretroviral drugs for HIV' and 'methadone' are associated with illnesses that are attributable to a patient's lifestyle, they score '0' on this criterion which lowers their overall rankings.[181]

When Cluster 6's criteria weights are used to rank the treatments, 'growth hormone for Prader-Willi Syndrome' is ranked 3rd which is the highest ranking for this treatment compared with the other clusters. This is because growth hormone is a treatment for children and relative to the other clusters, Cluster 6 has a high mean criterion weight for 'age'.

When the 'arbitrarily assigned' criteria weights are used to rank the treatments, some of the rankings noticeably change. For example, 'dialysis for renal disease' is ranked 9th whereas it is ranked between 1st and 5th using the cluster criteria weights.

Spearman's correlation coefficient measures the strength of association between two sets of ranked variables.[182] In this instance, Spearman's correlation coefficient between the ranked treatments using the random sample mean criteria weights and the ranked treatments using 'arbitrarily assigned' weights is 0.74, and it is 0.88 between the ranked treatments using the random sample mean criteria weights and the ranked treatments using 'evenly distributed' weights. Both sets of ranked treatments reveal a high positive association, which in large part is due to the way the treatments are categorised. However, the difference between the correlation coefficients is due to the differences in the criteria weights.

[181] The treatments are associated with a 'representative' patient. For example with HIV, the main cause of infection is men having unprotected sex with men. HIV can also be contracted by injecting or accidentally getting pricked with dirty needles or by infected mothers breastfeeding their babies. However, when categorising health treatments, it is the *main* reason for the illness or injury that is considered when deciding whether lifestyle has played a part in a person becoming ill.

[182] The Spearman correlation coefficient takes a value between -1 and + 1 with -1 indicating a perfect negative association between the variables and +1 indicating a perfect positive association between the variables.

Therefore the variation in health treatment rankings between groups is as a result of both the categorisation of the treatments and the differences in the criteria weights.

8.5 Conclusion

Cluster analysis was used to explore the heterogeneity of respondents' preferences. Random sample respondents were clustered into groups based on the similarity of their criteria weights. Chi-square testing reveals several associations between the clusters and the criteria weights (using two clusters and six clusters). However, as most of the statistically significant effects are small and the associations generally involve a small number of respondents, there does not appear to be specific patterns of preferences across the respondents. As discussed in the last chapter, it appears possible that the variation in respondents' preferences is related more to respondents' idiosyncrasies than to particular demographic characteristics.

However, to explore how differences in preferences affect the ranking of health treatments, the mean criteria weights from the clusters and the random sample are used together with 'evenly assigned' weights and 'arbitrarily assigned' weights to rank 14 health treatments. The exercise illustrates – assuming health treatments are appropriately categorised and the criteria weights accurately reflect respondent's preferences – that the rankings of health treatments will depend on whose criteria weights are used.

In the next chapter the criteria weights from the random sample are combined with other considerations within a priority-setting framework.

Appendix 8.2: Demographic characteristics of the respondents (two clusters)

Demographic Characteristics		Cluster 1 *n*=187		Cluster 2 *n*=135	
		No. in cluster	% of cluster	No. in cluster	% of cluster
Region	North Island	127	68%	103	76%
	South Island	60	32%	32	24%
Gender	Male	75	40%	55	41%
	Female	112	60%	80	60%
Age group	18-24	17	9%	6	4%
	25-34	17	9%	14	10%
	35-44	27	14%	15	11%
	45-54	41	22%	33	25%
	55-64	54	29%	34	25%
	65 yrs or over	31	17%	33	25%
Ethnicity	European	155	83%	120	89%
	Maori	7	4%	4	3%
	European/Maori	10	5%	6	4%
	Pacific Peoples	6	3%	1	1%
	Asian	7	4%	3	2%
	2 groups other	1	0.5%	1	1%
	Other	1	0.5%	0	0%
Qualification	No qualifications	11	6%	8	6%
	Secondary school	59	31%	48	36%
	Other post secondary school qualification	50	27%	33	24%
	University degree or equivalent	67	36%	46	34%
Employment	Employed full-time	74	39%	55	41%
	Employed part-time	32	17%	26	19%
	Looking for work	10	5%	3	2%
	Homemaker	14	8%	5	4%
	Student	14	8%	3	2%
	Retired	28	15%	35	26%
	Self-employed	10	5%	4	3%
	Other	5	3%	4	3%
Income group	$20,000 or less	20	11%	11	8%
	$20,001-$30,000	20	11%	13	9%
	$30,001-$50,000	26	13%	14	10%
	$50,001-$70,000	33	18%	28	21%
	$70,001-$100,000	26	14%	28	21%
	$100,001 or more	30	16%	24	18%
	Not given	32	17%	17	13%
Household composition	Couple/no children	72	38%	73	54%
	Parent(s) with child(ren)	69	37%	39	29%
	Extended family	16	9%	4	3%
	Alone	24	13%	15	11%
	Flatmate(s)	6	3%	3	2%
	Other	0	0%	1	1%

Demographic Characteristics		Cluster 1 *n*=187		Cluster 2 *n*=135	
		No. in cluster	% of cluster	No. in cluster	% of cluster
Health usage	Never	2	1%	0	0%
	Seldom	37	20%	21	16%
	Occasionally	98	52%	65	48%
	Frequently	50	27%	49	36%
Serious Illness	Yes	112	60%	99	73%
	No	75	40%	36	27%
Type of worker	Medical worker	10	5%	12	9%
	Health related worker	7	4%	4	3%
	Neither	170	91%	119	88%
Health insurance	Yes	78	42%	65	48%
	No	109	58%	70	52%

Appendix 8.2: Demographic characteristics of the respondents (six clusters)

Demographic Characteristics		Cluster 1 *n*=81		Cluster 2 *n*=33		Cluster 3 *n*=59		Cluster 4 *n*=77		Cluster 5 *n*=47		Cluster 6 *n*=25	
		No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster
Region	North Island	56	69%	27	82%	41	69%	57	74%	30	64%	19	76%
	South Island	25	31%	6	18%	18	31%	20	26%	17	36%	6	24%
Gender	Male	34	42%	13	39%	19	32%	32	42%	22	47%	10	40%
	Female	47	58%	20	61%	40	68%	45	58%	25	53%	15	60%
Age group	18-24	6	7%	1	3%	6	10%	2	3%	5	11%	3	12%
	25-34	10	12%	6	18%	1	2%	5	6%	6	13%	3	12%
	35-44	12	15%	3	9%	9	15%	8	10%	6	13%	4	16%
	45-54	23	28%	4	12%	8	14%	23	30%	10	21%	6	24%
	55-64	19	23%	9	27%	24	41%	20	26%	11	23%	5	20%
	65 yrs or over	11	14%	10	30%	11	19%	19	25%	9	19%	4	16%
Ethnicity	European	67	83%	31	94%	51	86%	69	90%	37	79%	20	80%
	Maori	4	5%	1	3%	2	3%	2	3%	1	2%	1	4%
	European/Maori	4	5%	1	3%	1	2%	4	5%	5	11%	1	4%
	Pacific Peoples	2	2%	0	0%	2	3%	1	1%	2	4%	0	0%
	Asian	2	2%	0	0%	3	5%	1	1%	2	4%	2	8%
	2 groups other	1	1%	0	0%	0	0%	0	0%	0	0%	1	4%
	Other	1	1%	0	0%	0	0%	0	0%	0	0%	0	0%
Qualification	No qualifications	2	2%	0	0%	4	7%	6	8%	5	11%	2	8%
	Secondary school	24	30%	10	30%	15	25%	25	32%	20	43%	13	52%
	Other post secondary school qualification	24	30%	8	24%	14	24%	20	26%	12	26%	5	20%
	University degree or equivalent	31	38%	15	45%	26	44%	26	34%	10	21%	5	20%

Demographic Characteristics		Cluster 1 *n*=81		Cluster 2 *n*=33		Cluster 3 *n*=59		Cluster 4 *n*=77		Cluster 5 *n*=47		Cluster 6 *n*=25	
		No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster
Employment	Employed full-time	36	44%	9	27%	19	32%	34	44%	19	40%	12	48%
	Employed part-time	16	20%	8	24%	10	17%	13	17%	6	13%	5	20%
	Looking for work	2	2%	1	3%	4	7%	2	3%	4	9%	0	0%
	Homemaker	6	7%	2	6%	4	7%	2	3%	4	9%	1	4%
	Student	6	7%	1	3%	5	8%	0	0%	3	6%	2	8%
	Retired	6	7%	10	30%	14	24%	20	26%	8	17%	5	20%
	Self-employed	5	6%	0	0%	2	3%	4	5%	3	6%	0	0%
	Other	4	5%	2	6%	1	2%	2	3%	0	0%	0	0%
Income group	$20,000 or less	8	10%	2	6%	8	14%	6	8%	4	9%	3	12%
	$20,001-$30,000	7	9%	3	9%	5	8%	5	6%	8	17%	5	20%
	$30,001-$50,000	7	9%	4	12%	9	15%	8	10%	10	21%	2	8%
	$50,001-$70,000	17	21%	5	15%	11	19%	17	22%	5	11%	6	24%
	$70,001-$100,000	12	15%	7	21%	5	8%	16	21%	9	19%	5	20%
	$100,001 or more	16	20%	7	21%	11	19%	14	18%	3	6%	3	12%
	Not given	14	17%	5	15%	10	17%	11	14%	8	17%	1	4%
Household composition	Couple/no children	34	42%	19	58%	19	32%	43	56%	19	40%	11	44%
	Parent(s) with child(ren)	29	36%	7	21%	20	34%	23	30%	20	43%	9	36%
	Extended family	5	6%	2	6%	7	12%	2	3%	4	9%	0	0%
	Alone	11	14%	4	12%	12	20%	8	10%	1	2%	3	12%
	Flatmate(s)	2	2%	1	3%	1	2%	0	0%	3	6%	2	8%
	Other	0	0%	0	0%	0	0%	1	1%	0	0%	0	0%

Demographic Characteristics		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
		81		33		59		77		47		25	
		No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster	No. in cluster	% of cluster
Health usage	Never	1	1%	0	0%	0	0%	0	0%	1	2%	0	0%
	Seldom	18	22%	5	15%	9	15%	13	17%	10	21%	3	12%
	Occasionally	46	57%	13	39%	33	56%	42	55%	19	40%	10	40%
	Frequently	16	20%	15	45%	17	29%	22	29%	17	36%	12	48%
Serious Illness	Yes	44	54%	27	82%	39	66%	53	69%	29	62%	19	76%
	No	37	46%	6	18%	20	34%	24	31%	18	38%	6	24%
Type of worker	Medical worker	5	6%	3	9%	5	8%	8	10%	0	0%	1	4%
	Health related worker	3	4%	3	9%	2	3%	0	0%	2	4%	1	4%
	Neither	73	90%	27	82%	52	88%	69	90%	45	96%	23	92%
Health insurance	Yes	32	40%	16	48%	30	51%	39	51%	16	34%	10	40%
	No	49	60%	17	52%	29	49%	38	49%	31	66%	15	60%

Appendix: 8.3: Results of chi-square tests (two clusters)

Independent Variables	Sig. $p < 0.05$ (2-tailed)	Cluster 1 (n = 187) Cluster 2 (n = 135)
Age (6)	No	
Age (3)	No	
Children/no children (2)		$\chi^2 = 6.270$, $p = 0.012$ $\chi^2 = 5.689$, $p = 0.017$ CV = 0.140, Res = -1.5
Ethnicity (3)	No	
Gender (2)	No	
Health insurance (2)	No	
Health usage (3)	No	
Qualifications (4)	No	
Qualifications (2)	No	
Regions (2)	No	
Regions (2) *Male*		$\chi^2 = 8.229$, $p = 0.004$ $\chi^2 = 7.130$, $p = 0.008$ CV = 0.252, Res = -1.9
Serious illness(2)		$\chi^2 = 6.270$, $p = 0.012$ $\chi^2 = 5.689$, $p = 0.017$ CV = 0.140, Res = -1.5
Serious illness (2) *Female*		$\chi^2 = 4.801$, $p = 0.028$ $\chi^2 = 4.147$, $p = 0.042$ CV = 0.158, Res = -1.4
Stated income (6)	No	
Stated income (3)	No	
Type of worker (2)	No	
Work status (6)	No	
Work status (3)		$\chi^2 = 12.457$, $p = 0.002$ $\chi^2 = 12.973$, $p = 0.002$ CV = 0.199, Res = -2.1*
Work status (3) *Female*		$\chi^2 = 13.663$, $p = 0.001$ $\chi^2 = 14.420$, $p = 0.001$ CV = 0.270, Res = −2.2*
Work status (2)	No	

Notes: The number of groups is in brackets
χ^2 Pearson's chi square (given first in italics), χ^2 Yate's continuity correction
CV: Cramer's V
*Significant standardised residual (highest residual from contingency table shown)
Effect size: small=0.01, medium=0.30, large=0.05

Appendix: 8.4: Results of chi-square tests (six clusters)

Independent Variables	Sig. $p < 0.05$ (2-tailed)	No. in cell	Cluster 1 $n = 81$	Cluster 2 $n = 33$	Cluster 3 $n = 59$	Cluster 4 $n = 77$	Cluster 5 $n = 47$	Cluster 6 $n = 25$
Age (6)	Gp 3	1 cell (8.3%) < 5			$n = 59$ $\chi^2 = 13.474$, $p = 0.019$ CV = 0.205 Res = −2.0*			
Age (3)	No							
Children/no children (2)	No							
Children/no children (2) *female*	Gps 2&5	0% < 5		$n = 17$ $\chi^2 = 5.587$, $p = 0.018$ $\chi^2 = 4.440$, $p = 0.035$ CV = 0.187, Res = 1.6			$n = 23$ $\chi^2 = 3.855$, $p = 0.050$ $\chi^2 = 3.021$, $p =$ **0.082** CV = 0.155, Res = 1.3	
Children/no children (2) *male*	Gp 3	0% < 5			$n = 17$ $\chi^2 = 4.400$, $p = 0.036$ $\chi^2 = 3.352$, $p =$ **0.067** CV = 0.197, Res = 1.5			
Ethnicity (3)	No							
Gender (2)	No							
Health insurance (2)	No							
Health insurance (2) *male*	Gp 1	0% < 5	$n = 34$ $\chi^2 = 5.178$, $p = 0.023$ $\chi^2 = 4.302$, $p = 0.038$ CV = 0.2, Res = 1.5					
Health usage (3)	Gp 1	0% < 5	$n = 80$ $\chi^2 = 6.153$, $p = 0.048$ CV = 0.139, Res = −1.8					
Qualifications (4)	No							
Qualifications (2)	Gps 5&6	0% < 5					$n = 47$ $\chi^2 = 4.568$, $p = 0.033$ $\chi^2 = 3.903$, $p = 0.048$ CV = 0.119, Res = 1.5	$n = 25$ $\chi^2 = 4.956$, $p = 0.026$ $\chi^2 = 4.052$, $p = 0.044$ CV = 0.124, Res = 1.7
Qualifications (2) *Female*	Gp 6	0% < 5						$n = 15$ $\chi^2 = 7.028$, $p = 0.005$ $\chi^2 = 6.543$, $p = 0.011$ CV = 0.204, Res = 2.1*
Qualifications (2) *Male*	Gp 5	0% < 5					$n = 22$ $\chi^2 = 4.381$, $p = 0.036$ $\chi^2 = 3.436$, $p =$ **0.064** CV = 0.184, Res = 1.5	

Independent Variables	Sig. $p < 0.05$ (2-tailed)	No. in cell	Cluster 1 $n = 81$	Cluster 2 $n = 33$	Cluster 3 $n = 59$	Cluster 4 $n = 77$	Cluster 5 $n = 47$	Cluster 6 $n = 25$
Regions (2)	No							
Regions (2) *Male*	Gps 4&5	0% < 5				$n = 32$ $\chi^2 = 4.893$, $p = 0.027$ $\chi^2 = 3.938$, $p = 0.047$ CV = 0.194, Res = 1.6	$n = 22$ $\chi^2 = 6.581$, $p = 0.010$ $\chi^2 = 5.309$, $p = 0.021$ CV = 0.225, Res = 2.0*	
Serious illness(2)	Gps 1&2	0% < 5	$n = 81$ $\chi^2 = 6.017$, $p = 0.014$ $\chi^2 = 5.373$, $p = 0.020$ CV = 0.137, Res 1.7	$n = 33$ $\chi^2 = 4.319$, $p = 0.038$ $\chi^2 = 3.553$, $p =$ **0.059** CV = 0.116, Res = 1.6				
Serious illness (2) *Female*	Gp 1	0% < 5	$n = 47$ $\chi^2 = 4.664$, $p = 0.031$ $\chi^2 = 3.929$, $p = 0.047$ CV = 0.156, Res 1.5					
Stated income (6)	Gp 5	2 cells (16.7%) < 5					$n = 39$ $\chi^2 = 11.928$, $p = 0.036$ CV = 0.209, Res 1.8	
Stated income (3)	No							
Type of worker (2)	No							
Work status (6)	No							
Work status (3)	Gps 1&4	0% < 5	$n = 77$ $\chi^2 = 9.667$, $p = 0.008$ CV = 0.176, Res = −2.4*			$n = 75$ $\chi^2 = 9.027$, $p = 0.011$ CV = 0.17, Res = −2.3*		
Work status (3) *Female*	Gps 1&4	0% < 5	$n = 45$ $\chi^2 = 10.531$, $p = 0.005$ CV = 0.237, Res = −2.5*			$n = 45$ $\chi^2 = 7.687$, $p = 0.021$ CV = 0.202 Res = −2.0*		
Work status (2)	Gp 1	0% < 5	$n = 77$ $\chi^2 = 4.276$, $p = 0.039$ $\chi^2 = 3.728$, $p =$ **0.053** CV = 0.117, Res = −1.4					
Work status (2) *Female*	Gps 1&3	0% < 5	$n = 45$ $\chi^2 = 6.082$, $p = 0.014$ $\chi^2 = 5.243$, $p = 0.022$ CV = 0.180, Res = −1.7		$n = 39$ $\chi^2 = 5.414$, $p = 0.020$ $\chi^2 = 4.585$, $p = 0.032$ CV = 0.17, Res = 1.6			

Notes: The number of groups is in brackets
χ^2 Pearson's chi square (given first in italics), χ^2 Yate's continuity correction
CV: Cramer's V
*Significant standardised residual (highest residual from contingency table shown)
Effect size: small=0.01, medium=0.30, large=0.05

~ Chapter 9 ~

Prioritisation framework

9.1 Introduction

In this chapter, the criteria weights, along with other relevant considerations for prioritising health technologies such as 'cost' and 'quality/strength of medical evidence' are brought together within the MCDA framework. The chapter begins with a discussion of some of the additional considerations that were raised in the focus groups or mentioned in the literature, but were deliberately not included as criteria in the decision survey. The 14 health treatments from the focus groups are then used in two examples to illustrate how decision-makers can allocate funding over a range of health treatments taking into account the criteria weights and the other considerations.

9.2 Including additional considerations in the prioritisation framework

> "A hip replacement is easier to cure whereas the methadone treatment for drug addicts takes a long time."
>
> *Member of the retiree's focus group*

> "I think it's important to distinguish between possible or likely efficacy and effectiveness and proven efficacy and effectiveness."
>
> *Member of the public health group*

As discussed in Chapter 4, several considerations raised by focus groups participants or mentioned in the literature such as 'cost', 'duration of treatment', 'number of patients' and 'strength of medical evidence' were not included as criteria in the decision survey but will be included as 'additional considerations' within the MCDA framework.

'Cost', as explained in Chapter 4, was not included as a criterion in the decision survey for the following reasons. As respondents were asked to choose between two hypothetical *patients* and not two *health programmes*, the relevant cost is 'cost per patient'. However, including 'cost per patient' in the survey is problematic for two reasons. First, it is unfeasible for some

treatments to be at the 'cost per patient' level. For example, vaccines are purchased at a programme level and not at an individual patient level. Second, 'cost per patient' adds complexity to the trade-off questions and also creates uncertainty relating to opportunity cost (Bryan et al. 2002). For example, when the lower cost treatment is chosen out of two treatment options, it can mean one of three things: 1) that money has been saved, 2) that more of the lower cost treatment can be purchased, or 3) that the other treatment option is more valuable.

'Duration of treatment', 'number of patients' and 'strength of medical evidence' were also suggested as potential criteria by focus group participants. Once again, the main reason for not including these considerations as criteria in the decision survey is because respondents were asked to choose between two hypothetical *patients* and not between two *treatment programmes*. In addition, a criterion such as 'strength of medical evidence' is difficult for some respondents to understand, which leads to trade-off questions being inconsistently interpreted. However, these additional factors will still be considered in the priority-setting process. Before this process is outlined, the additional considerations are explained further.

As can be seen in Table 9.1, some 'other considerations' have been added to the rankings table: 'total cost', 'cost per patient', 'number of patients', 'duration of treatment' and 'strength of medical evidence'.[183]

[183] As discussed in Chapter three, the data for the costs, number of patients affected per year, the duration of treatment and strength of medical evidence were gathered from the literature and advice from health experts. Although the data are as accurate as possible, they are estimates only.

Table 9.1: Ranking table including additional considerations

	CRITERIA								OTHER CONSIDERATIONS				
ALTERNATIVE click to open	Treatment options for this patient	Age of patient	Benefit to others (eg family or society)	Patient's health before treatment	Benefit to patient (ie length and/or quality of life)	Illness or injury caused mainly by lifestyle choices	RANK	TOTAL SCORE	Total cost ($ million)	Cost per patient ($)	Number of patients	Duration of treatment	Quality/strength of medical evidence
Hip replacements	this is the ONLY treatment available	15-64 years	small	poor (but not immediately life threatening)	large	no	1st	66.9 %	119	17000	7000	one-off	strong
Dialysis	this is the ONLY treatment available	65+ years	small	will die soon without treatment	large	yes	2nd	61.0 %	22	50000	440	5-10yrs	strong
Leukaemia drugs	this is the best treatment (there are less effective alternatives)	15-64 years	small	will die soon without treatment	medium	no	3rd	60.7 %	2.4	60000	40	5-10yrs	strong
HIV drugs	this is the ONLY treatment available	15-64 years	small	poor (but not immediately life threatening)	large	yes	4th	54.1 %	1.6875	13500	125	life	strong
Rheumatoid arthritis drugs	this is the best treatment (there are less effective alternatives)	15-64 years	small	poor (but not immediately life threatening)	medium	no	5th	46.5 %	0.09	3000	30	10+yrs	strong
Growth hormone	this is the ONLY treatment available	0-14 years	small	fair (neither good nor bad)	small	no	6th	44.2 %	0.414	18000	23	10+yrs	strong
Hand sanitiser^	this is the best treatment (there are less effective alternatives)	0-14 years	large	relatively good (though treatment is still beneficial)	small	no	7th	39.0 %	0.136	0.34	400000	one-off	weak
Postnatal depression service	this is the best treatment (there are less effective alternatives)	15-64 years	large	fair (neither good nor bad)	small	no	8th	39.0 %	1	400	2500	one-off	strong
Methadone	this is the best treatment (there are less effective alternatives)	15-64 years	large	poor (but not immediately life threatening)	small	yes	9th	33.6 %	20	5000	4000	5yrs	strong
PET Scan	this is the best treatment (there are less effective alternatives)	15-64 years	small	fair (neither good nor bad)	small	no	10th	26.9 %	8.5	1700	5000	one-off	strong
Cervical cancer vaccine^	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	20.2 %	16	1600	10000	5yrs	weak
Erectile dysfunction drugs	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	20.2 %	1.422	158	9000	one-off	strong
IVF	this is the best treatment (there are less effective alternatives)	15-64 years	small	relatively good (though treatment is still beneficial)	small	no	11th=	20.2 %	17.64	14400	1225	one-off	strong
Statins	this is the best treatment (there are	15-64 years	small	fair (neither good nor bad)	small	yes	14th	14.1 %	11	50	220000	life	strong

Cost is included in two ways: 'cost per patient' and 'total cost ($millions)' which is the number of patients being treated over a one year period multiplied by cost per patient. The 'number of patients' refers to how many patients will be treated over a one-year period. 'Duration of treatment' refers to whether a treatment is a 'one-off' such as hip replacements or whether treatments are ongoing – for example, growth hormone treatment is given to children until they reach adulthood and the cervical cancer vaccine is a five-year programme.

If patients are receiving a benefit from treatment and they change to a more effective treatment, then the *incremental* benefits and costs of treatment should be considered and not the *total* benefits and costs. A 'marginal' approach ensures that the opportunity costs of new and existing treatments are made explicit (Grocott 2009). By comparing the marginal benefits and marginal costs of existing services with those of proposed new technologies, resources can be allocated or re-allocated as efficiently as possible within a defined budget (Mitton & Donaldson 2003). Therefore, ideally, the marginal benefits and marginal costs should be used in a prioritisation framework. However, the data used in this part of the thesis, is for illustrative purposes only, i.e. to demonstrate how the proposed prioritisation framework could be applied in practice. For simplicity, and because information relating to the marginal benefits and marginal costs was not available for all the treatments, the *total* benefits and costs of treatment are used. It is assumed that patients are either not receiving treatment or if they are receiving treatment, that they are *not* receiving any benefit.

9.2.1 Quality/strength of medical evidence

'Quality/strength of medical evidence' is an important element in a priority-setting process and relates to the safety, efficacy and effectiveness of health treatments. There are a number of different rating systems used to assess the strength of medical evidence (Treadwell et al. 2006). Mitton & Patten (2004) conducted a research project in a single health authority in Alberta, Canada. Decision-makers in the project suggested "using a mix of 'soft' and 'hard' forms of evidence in priority-setting" (p 146): when 'hard' evidence is not available, such as randomised controlled trials,[184] funding decisions can be made using 'soft' evidence such as non-randomised quantitative studies, anecdotal reports and expert opinion.

[184] A randomised control trial is a study in which the efficacy or effectiveness of a specific drug or treatment is tested. Study participants are randomly allocated into two or more groups with one of the groups being a 'control' group where participants receive a placebo or no intervention at all. The outcomes of each group are then compared to assess the impact of the drug or treatment.

The GRADE[185] approach is a formal system for grading health treatments based on medical evidence (GRADE working group 2004). GRADE separates 'quality of evidence' from 'strength of recommendations'. 'Quality of evidence' refers to the advantages and disadvantages of treatment and is evaluated on four levels – high, moderate, low and very low. 'Strength of recommendation' refers to how much confidence there is in the quality of evidence and is evaluated on two levels – strong and weak. These explicit criteria ensure that decisions relating to medical evidence are transparent (Guyatt et al. 2008).

Treadwell et al. (2006) describes a new points system for rating the strength of evidence. The authors distinguish between quantitative and qualitative aspects, use *a priori* criteria for judgements, and include the direct impact of meta-analysis[186] and sensitivity analysis on the ratings.

Mullen (2004) reviewed a large number of priority-setting projects to analyse the main issues associated with priority setting in health care, including the elicitation of relevant criteria. The author found that the "rise of evidence-based medicine" (p 54) has led to many projects including 'strength or quality of evidence' as a separate criterion. In this thesis, quality/strength of medical evidence is not a separate criterion but is considered in addition to the six criteria. As previously mentioned, because some members of the general public may have had difficulty understanding the criterion 'quality/strength of medical evidence', and interpreting the two criterion levels 'strong' and 'weak', 'quality/strength of medical evidence' is considered separately.

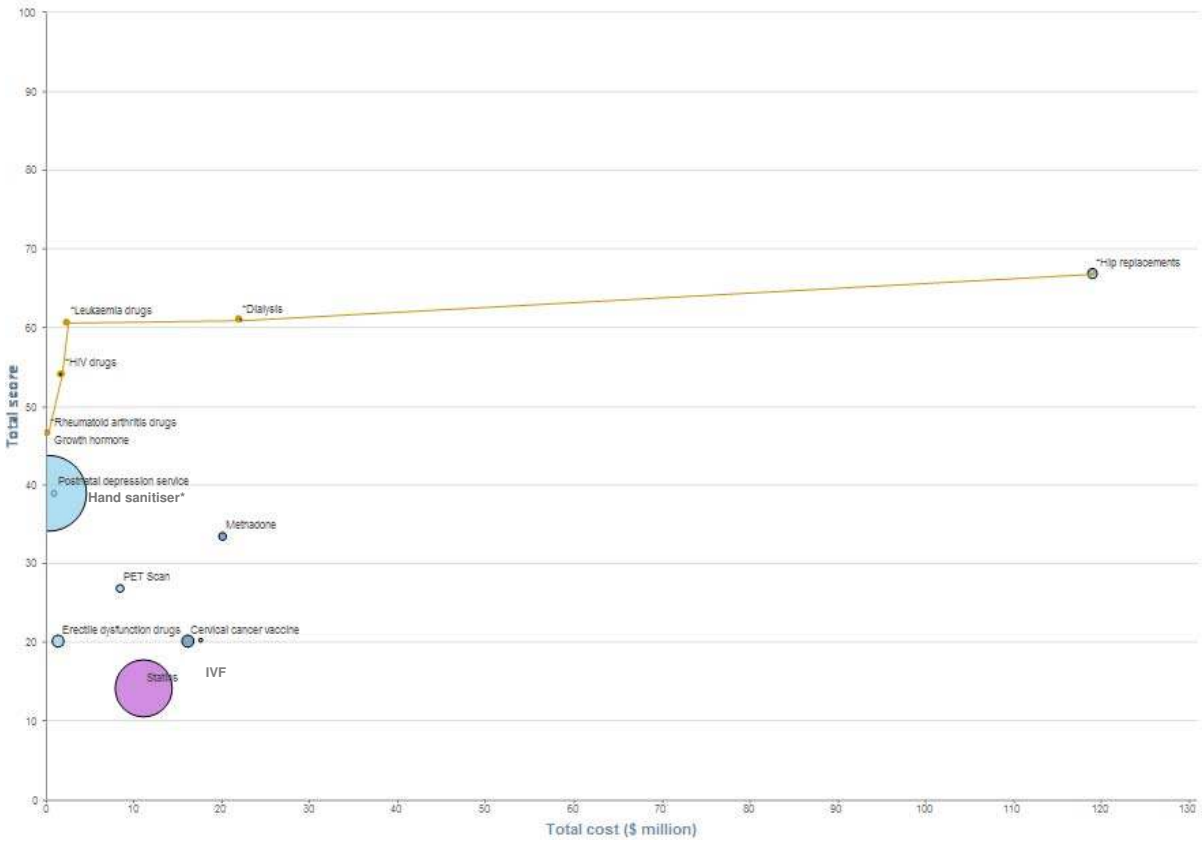
9.3 Value for money chart

> "On a personal level it is tragic [not funding high cost treatments] but on a society level somebody has to make the really tough decisions. As a parent if you were told that your son or daughter's medicine was going to cost $4m a year what would you say?"
>
> *Member of the GP practice focus group*

[185] GRADE is an acronym for Grades of Recommendation, Assessment, Development and Evaluation. This system is used globally by more than 25 organisations including the World Health Organisation, the American College of Physicians and the Cochrane Collaboration (Guyatt 2008).
[186] A meta-analysis combines the results of a number of similar studies.

The information in Table 9.1 can be used to prioritise health treatments. For example, treatments can be ranked according to their total score or according to 'other considerations' such as total cost, or treatments could be ranked based on a combination of considerations. To make it easier for decision-makers to consider more than one variable at a time, 1000Minds software produces a Value for Money (VfM) chart which brings together all the variables needed to select and prioritise health treatments (Golan & Hansen 2008). For example, in Figure 9.1 the 14 health treatments from the focus groups are plotted in a VfM chart using five variables from Table 9.1 – 'total score', 'total cost', 'number of patients', 'duration of treatment' and 'quality/strength of medical evidence'.[187]

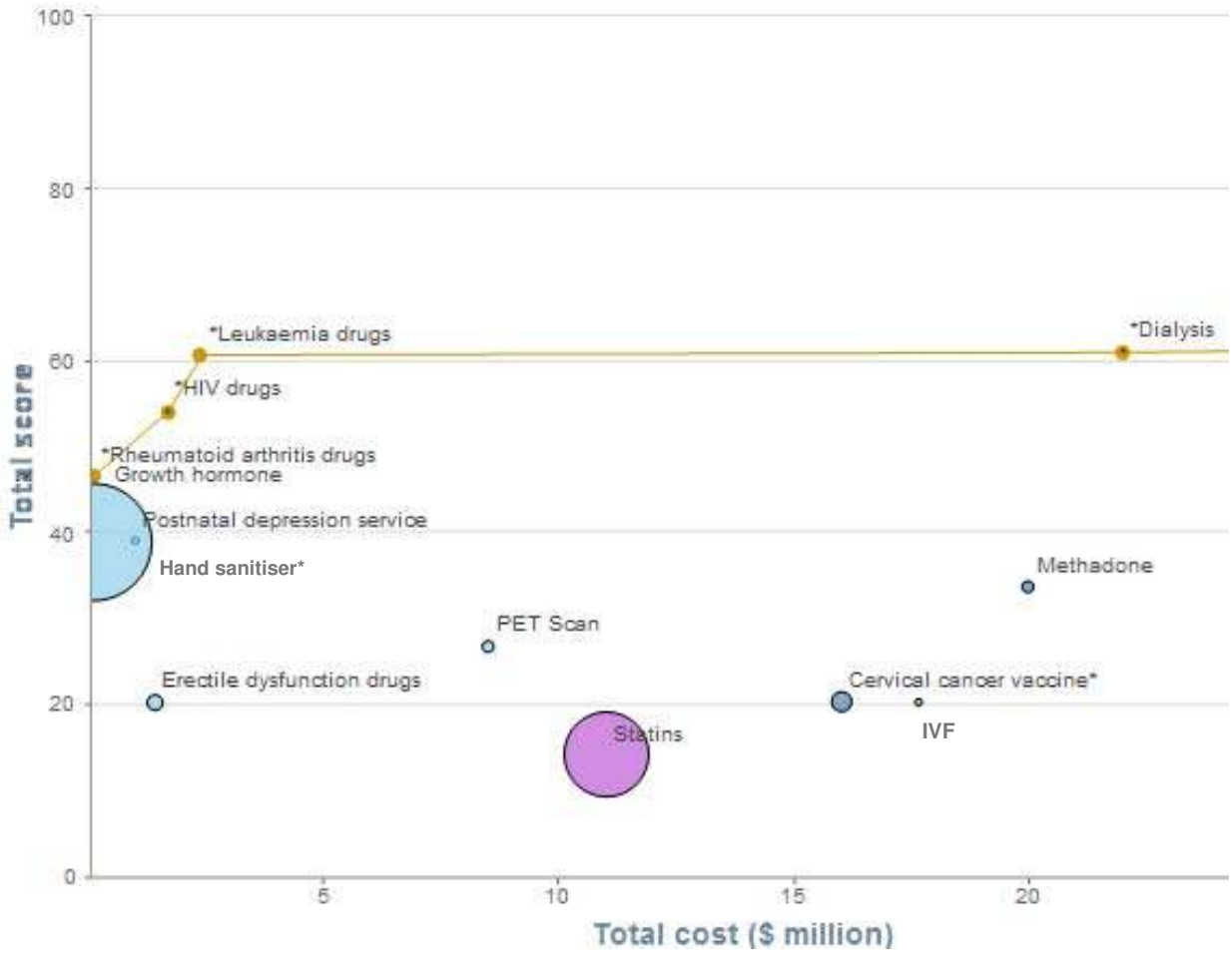
Figure 9.1: Value for money chart



To get a closer look at the VfMChart, Figure 9.2 displays the VfM Chart without the outlying treatment 'hip replacements'.

[187] Although the 14 health treatments are used for illustrative purposes, the actual data are real. The health treatments have been categorised as accurately as possible on the six criteria and the data for the costs, number of patients, duration of treatment and strength of medical evidence were gathered from the literature and advice from health experts.

Figure 9.2: Value for money chart (13 treatments)



The vertical axis displays each health treatment's 'total score' which in essence is the 'total benefit' as the total score is obtained by categorising and scoring each treatment according to the six criteria and summing the criteria weights. The horizontal axis displays the total cost of treatment (i.e. cost of treatment × number of patients) for a one year period. The size of a bubble represents the number of patients receiving treatment – the larger the bubble, the greater the number of patients receiving treatment. For example, 'hand sanitiser' is represented by a relatively large bubble because it affects 400,000 children whereas 'leukaemia drugs' is represented by a relatively small bubble as it affects only 40 patients. The colour of a bubble represents the duration of treatment. For instance, treatments that *potentially* need to be funded for the life of the patient are coloured purple whereas one-off treatments are coloured light blue. Quality/strength of medical evidence is indicated by an

asterisk. The 'cervical cancer vaccine' and 'hand sanitiser' both have asterisks indicating that the quality/strength of medical evidence relating to these treatments is weak.[188]

Alternative representations are possible. How the data are presented depends on the availability and the format of the data. For instance, in Figure 9.1 the total scores for each treatment are determined by the criteria weights from the random sample and are plotted against 'cost per patient'; in Section 9.5 an example is given where the total scores for each treatment are determined by the criteria weights from the health services researchers' sample. Similarly, the size and the colour of the bubbles can represent any variable. For example, additional considerations could include ethical, legal, strategic or political considerations, or the robustness of the estimated costs.

How much 'weight' is put on the additional considerations such as 'quality/strength of evidence' is up to the decision-makers. For example, a treatment such as the cervical cancer vaccine has weak evidence to support the clinical benefits, but as it is a one-off programme and is relatively cost-effective, decision-makers may decide to fund this treatment before funding another treatment that has stronger medical evidence.

Referring back to Figure 9.2, treatments closest to the vertical axis have the lowest cost per patient and treatments farthest away from the horizontal axis have the highest benefit. Therefore, ignoring for now the other considerations (number of patients, duration of treatment and quality/strength of medical evidence), the treatments with the highest scores and the lowest costs are preferred to other treatments. The 'Pareto (efficiency) frontier', which is the brown line in Figures 9.1 and 9.2, identifies the 'dominant' treatments. All else being equal, there are no other treatments that have both lower cost per patient *and* a higher total score than these dominant treatments. If a line were to be drawn from the origin of the chart to each treatment, the gradient of the line would indicate the 'efficiency' of the treatment – a comparatively steep line indicates a comparatively high benefits/total cost ratio.

As can be seen in Figures 9.1 and 9.2, the 'Pareto frontier' includes 'rheumatoid arthritis drugs' 'HIV drugs', 'leukaemia drugs', 'dialysis' and 'hip replacements.

[188] As both the 'cervical cancer vaccine' and the 'hand sanitiser' are relatively new treatments, the strength of medical evidence available for these treatments is relatively weak. When health treatments are relatively new, particularly those used to treat only a small number of the population, it is often not possible to obtain comparable information on the safety, efficacy and quality of the treatments.

The VfM chart presents the prioritisation variables in a way that is easy for the decision-makers to interpret. However, the budget allocation table which accompanies the VfM chart and includes all the variables needed to prioritise treatments, is also useful for tracking decisions and assessing the impact on the *overall* budget. For example, in Table 9.2, the budget allocation table includes all the variables from the rankings table (Table 9.1) as well as additional variables such as the cost/benefit ratio and the option to enter a budget constraint.[189]

Table 9.2: Budget allocation table

✓	✘	?	Technology	Cost per patient ($)	Total score	No. of patients	Cost/ Benefit Ratio	Total cost ($ million)	Duration of treatment	Quality/ strength of medical evidence	Rank
✓			Hip replacements	17000	66.9%	7000	254.018	119	one-off	strong	1st
✓			Dialysis	50000	61.0%	440	820.151	22	5-10yrs	strong	2nd
✓			Leukaemia drugs	60000	60.7%	40	987.827	2.4	5-10yrs	strong	3rd
✓			HIV drugs	13500	54.1%	125	249.384	1.6875	life	strong	4th
✓			Rheumatoid arthritis drugs	3000	46.5%	30	64.5112	0.09	10+yrs	strong	5th
		?	Growth hormone	18000	44.2%	23	407.041	0.414	10+yrs	strong	6th
		?	Hand sanitiser*	0.34	39.0%	400000	0.00871	0.136	one-off	weak	7th
		?	Postnatal depression service	400	39.0%	2500	10.2568	1	one-off	strong	8th
		?	Methadone	5000	33.6%	4000	148.656	20	5yrs	strong	9th
		?	PET Scan	1700	26.9%	5000	63.1695	8.5	one-off	strong	10th
	✘		Cervical cancer vaccine*	1600	20.2%	10000	79.2244	16	5yrs	weak	11th=
	✘		Erectile dysfunction drugs	158	20.2%	9000	7.82341	1.422	one-off	strong	11th=
	✘		IVF	14400	20.2%	1225	713.02	17.64	one-off	strong	11th=
		?	Statins	50	14.1%	220000	3.54086	11	life	strong	14th

✓ Select
✘ Reject
? Undecided

Total selected	145.18
Total rejected	35.06
Total undecided	41.05
TOTAL COST (of purchasing all treatments)	**$221.29m**

Budget/constraint (optional):	
No yet allocated:	

Decision-makers can use the table to 'select' various health treatments and assess the impact on the budget. For example, as shown in Table 9.2, when the five health treatments on the Pareto frontier are 'selected', 'cervical cancer vaccine', 'erectile dysfunction drugs' and 'IVF' 'rejected', and the remaining five treatments 'undecided', the total cost of the five 'selected' treatments is $145,177,500, the cost of the 'rejected' treatments is $35,062,000 and the cost of

[189] The software automatically updates the budget allocation table and the VfM chart when any changes are made to the rankings table.

the 'undecided' treatments is $41,050,000. If all 14 of the treatments were to be funded, the total cost would be $221,289,500.

A factor to consider when 'selecting' treatments is whether the treatment can be funded at an individual level or whether the entire health programme needs to be purchased. For example, the cervical cancer vaccination programme is a five year programme in which 50,000 young women will be vaccinated. Although 'cost per patient' for the vaccine is included in the table, the *entire* programme needs to be funded. Treatments that are divisible could be highlighted (in bold for example) to make decision-makers aware of which treatments need to be purchased at the programme level and which treatments can be purchased at the individual level, in which case the number of patients could be altered.[190]

The VfM chart and table enable decision-makers to compare alternative combinations of treatments to arrive at an 'optimal' portfolio of treatments (Golan & Hansen 2008). For example, decision-makers can allocate a fixed budget across a full range of health services or specific health areas by entering an amount in the 'budget/constraint' box at the bottom of the table and 'selecting' or 'deselecting' treatments to fit within the allocated budget.[191] Decision-makers may also wish to re-allocate resources when there has been a cut in funding, or to compare the current year's funding with the previous year's funding in terms of the total cost, number of patients treated, treatments purchased for instance.

An example of a VfM chart and budget allocation table using QALYs is presented in the next section.

9.4 Prioritising health treatments using QALYs

Many health funding agencies, including PHARMAC in New Zealand and NICE in the UK, use QALYs when prioritising health treatments.[192] In this section, the health treatments from the focus groups and the criteria weights from the health services researchers' sample are used

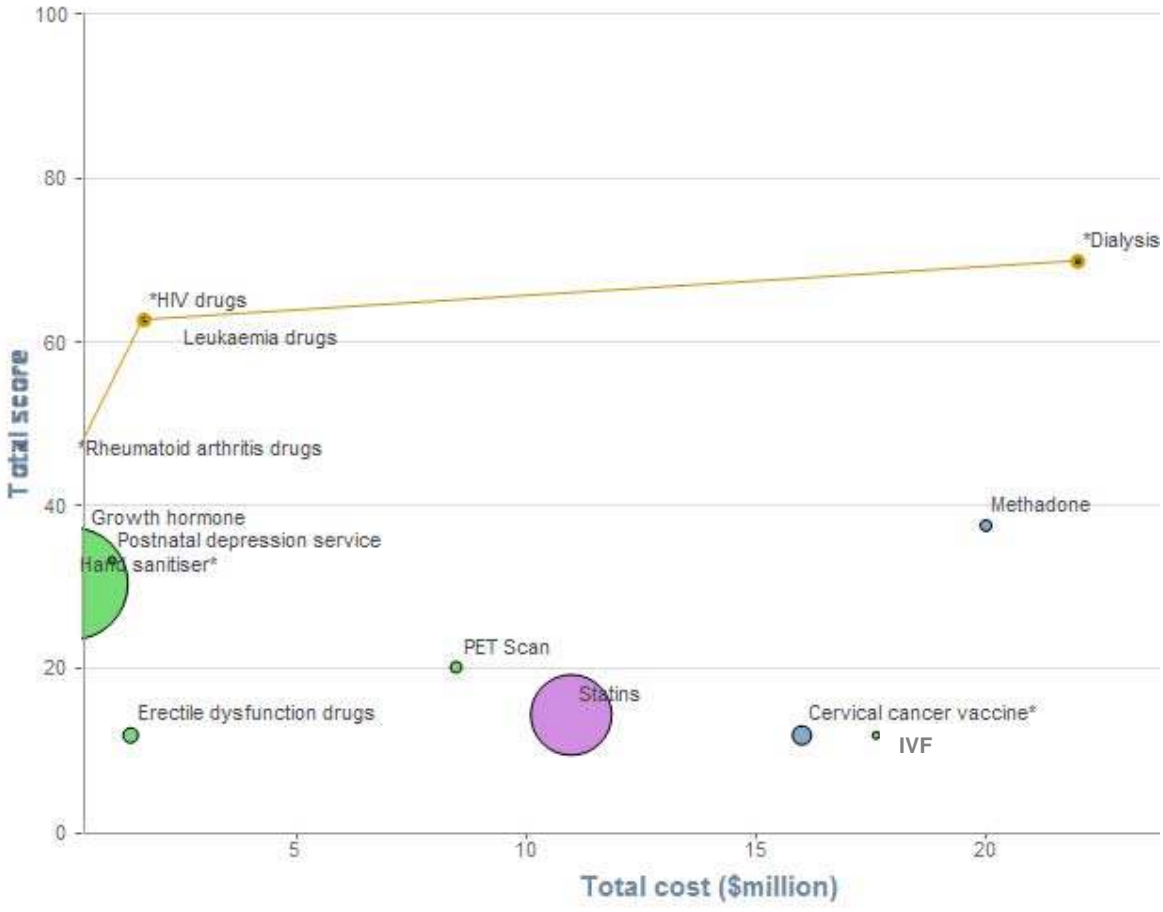
[190] Some treatments may have a relatively low 'cost per patient' but if the minimum number of people who require treatment is large, the total cost could be prohibitive. On the other hand, a treatment that has a relatively high 'cost per patient' may only affect a small number of patients and therefore the total cost is relatively small.
[191] This type of cost analysis could also be done using an excel spreadsheet. The advantage of using the budget allocation table in 1000Minds is that the table and VfM chart are automatically updated when any changes are made to the rankings table, including changes in the criteria weights.
[192] As discussed previously, factors other than QALYs are considered by PHARMAC and NICE when prioritising health treatments. However, the way this is done is often not made explicit.

to illustrate how QALYs, in combination with 'other considerations' can be used *explicitly* to prioritise health treatments.

In Figure 9.3, 13 of the 14 health treatments are displayed in a VfM chart, with 'total score' on the vertical axis and 'total cost' on the horizontal axis.[193] 'Total score' includes the criteria weights for the six criteria – 'treatment options for this patient', 'age of patient', 'benefit to others', 'patient's health before treatment', 'illness or injury caused mainly by lifestyle choices' and 'benefit to patient' (in terms of QALYs) – from the health services researchers' decision survey. The size of the bubbles represents the number of patients, the colour of the bubbles represents duration of the treatment and an asterisk identifies whether the treatment has strong or weak quality/strength of medical evidence.

As only 12 health services researchers completed the survey, the 'total score' for each treatment is a rough estimate for illustrative purposes only.

Figure 9.3: Value for money chart using QALYs



[193] To get a closer look at the chart, one 'outlying' treatment (hip replacements) is not displayed.

The relevant prioritisation variables are listed in the budget allocation table in Table 9.3. Another possible consideration for funding agencies such as PHARMAC is whether a treatment needs to be purchased as part of a bundle or whether it can be purchased separately.

Table 9.3: Budget allocation table (QALYS)

✓	✘	?	Technology	QoL gains	Total score	No. of patients	Total cost ($ million)	Duration of treatment	Quality/ strength of medical evidence	Rank
✓			Dialysis	large	69.8%	440	22	5-10yrs	strong	1st
		?	Hip replacements	large	68.4%	7000	119	one-off	strong	2nd
✓			HIV drugs	large	62.6%	125	1.6875	life	strong	3rd
		?	Leukaemia drugs	medium	57.8%	40	2.4	5-10yrs	strong	4th
✓			Rheumatoid arthritis drugs	medium	44.5%	30	0.09	10+yrs	strong	5th
		?	Methadone	small	37.4%	4000	20	5yrs	strong	6th
		?	Growth hormone	small	35.7%	23	0.414	10+yrs	strong	7th
		?	Postnatal depression service	small	33.2%	2500	1	one-off	strong	8th
		?	Hand sanitiser*	small	30.2%	400000	0.136	one-off	weak	9th
		?	PET Scan	small	20.1%	5000	8.5	one-off	strong	10th
		?	Statins	small	14.3%	220000	11	life	strong	11th
	✘		IVF	small	11.9%	1225	17.64	one-off	strong	12th=
	✘		Cervical cancer vaccine*	small	11.9%	10000	16	5yrs	weak	12th=
		?	Erectile dysfunction drugs	small	11.9%	9000	1.422	one-off	strong	12th=

✓ Select
✘ Reject
? Undecided

Total selected		23.78
Total rejected		33.64
Total undecided		163.87
TOTAL COST (of purchasing all treatments)		**221.29**

Budget/constraint (optional):	
No yet allocated:	

9.5 Discussion

The treatments on the Pareto frontier in Table 9.2 are different from the treatments on the Pareto frontier in Table 9.3. This is because the criteria weights from the random sample are used to obtain the total benefit scores in Table 9.2 and the criteria weights from the health services researchers' sample are used to obtain the benefit scores in Table 9.3. In both Tables 1 and 2, the total scores are plotted against 'total cost'. Though it is possible to plot the total scores against 'cost per patient', this can only be done if the treatments are divisible (e.g. hip replacements are purchased separately whereas the cervical cancer vaccine is purchased as an entire programme). In addition, 'cost per patient' assumes that there are constant returns to scale (Golan & Hansen forthcoming). Further, as affordability is an important factor in

allocating a budget, it is the *total* cost of a programme that ought to be considered (Golan & Hansen forthcoming). For example, hand sanitiser is only 34c per child over four months but if 400,000 children use the hand sanitiser, the total cost of $136,000 may be considered unaffordable.

There are many practical uses for the VfM chart and associated budget allocation table, including its use as a communication tool. The outcome of accepting and/or rejecting particular treatments can be analysed in an easy-to-use format. The VfM chart and budget allocation table enable decision-makers to allocate fixed budgets across a range of services, to keep track of previous decisions and to re-allocate resources if the budget has been cut while taking into account all relevant considerations.

9.6 Conclusion

In the final step of MCDA, the criteria weights and other considerations are used to *assist* decision-makers in priority-setting. In this chapter, a VfM chart, which brings together all the prioritisation variables needed to select and prioritise health treatments, is used together with a budget allocation table, to illustrate how health treatments can be prioritised. Ultimately it is up to the decision-makers to choose which treatments to fund within an available budget, but if decisions are made explicitly within a transparent and robust framework that includes all relevant considerations (including the preferences of key stakeholders) then there is likely to be more ‘buy-in’ and acceptance in the outcome.

~ Chapter 10 ~

Conclusion

10.1 Introduction

New Zealand is one of many countries grappling with the issue of how to allocate limited resources across a range of health and disability services at a time when demand for health care is increasing and costs are escalating. In this thesis I use MCDA to develop a framework which can be used by decision-makers to prioritise publicly-funded health care. This chapter begins with an overview of the thesis, followed by a discussion of the key findings, and concludes with the limitations of the thesis and suggestions for future research.

10.2 Overview of the thesis

A typical MCDA process begins with identifying what needs to be prioritised (in this thesis it is health care treatments and/or programmes) and deciding who should be involved in the priority-setting process. The next stage is to establish the criteria by which the health treatments will be prioritised and to determine their relative importance. Finally, the criteria weights (which represent the relative importance of each of the criteria) are used to assist decision-makers in priority-setting.

To establish the relevant criteria by which health treatments can be prioritised, suggested criteria from six focus groups were amalgamated with advice from health experts and potential criteria from the literature. Six criteria were established. The relative importance of the criteria was determined using a choice-based online survey implemented through 1000Minds software (Ombler & Hansen 2012) and the PAPRIKA scoring method (Hansen & Ombler 2008). Three groups of respondents completed the survey – a random sample from the New Zealand electoral roll, a 'snowball' sample and a group of health services researchers.

Regression analysis was used to determine whether the demographic characteristics of respondents can predict preferences, and cluster analysis was used to explore the heterogeneity of respondents' preferences. (The main findings are presented in the next section.)

The criteria weights from the random sample were then brought together with cost and other 'additional factors' in a prioritisation framework. A value for money (VfM) chart and associated budget allocation table were used to illustrate how health treatments can be prioritised by decision-makers in a clear, consistent and transparent way.

10.3 Key findings

Key findings from the thesis can be divided into two main groups: methodology and results.

10.3.1 Methodology

Using 1000Minds software and the PAPRIKA scoring method to elicit the preferences of respondents is effective. The surveys[194] are straightforward to create, cost-effective to implement and easy to administer. In terms of the survey format (i.e. the survey design and instructions) 81.8% of the snowball sample and 84.5% of the random sample found the survey format either easy, or very easy, to follow. In regard to the survey's ability to accurately capture respondents' preferences, the results of a 'test re-test' (discussed in Chapter 5) indicate that the decision survey consistently measured the preferences of respondents.

A useful feature of 1000Minds software is the VfM chart and associated budget allocation table that enable decision-makers to consider all relevant prioritisation variables in a transparent and consistent way. Changes can be made to any of the variables, including the benefit scores, and the software automatically updates the VfM chart and table.

In contrast to other methods where the criteria weights are calculated for the overall sample, the PAPRIKA method produces a set of criteria weights for *each* individual. This allows the criteria weights to be analysed for the overall sample, for sub-groups and for individuals.

10.3.2 The relative importance of prioritisation criteria to New Zealanders

The mean criteria weights for the 322 random sample respondents are presented in Table 10.1. The criteria weights represent how important each criterion is in relation to each other. For example, on average, the criterion 'need' (28.4%) is twice as important as the criterion 'age' (14.2%).

[194] A ranking survey and a decision survey were used in this thesis.

Table 10.1: Random sample criteria weights

RANDOM SAMPLE (*N*=322)	Criteria weights
Need *Patient's health before treatment (health status)*	28.4%
Individual benefit *Benefit to patient (ie length and/or quality of life)*	22.0%
Age *Age of patient*	14.2%
Lifestyle *Illness or injury NOT caused by lifestyle choices*	12.8%
Societal benefit *Benefit to others (eg family or society)*	12.1%
No alternative treatment *Only treatment option available for this patient*	10.5%

The two most important criteria are 'need' and 'individual benefit' which, as discussed in Chapter 6, is in line with studies of a similar nature. As can be seen in Table 10.1, 'age' and 'lifestyle' are considered to be more important that 'societal benefit' and 'no alternative treatment'. The age and lifestyle of a patient are not often considered in prioritisation processes (usually for anti-discriminatory reasons). Therefore there are few studies where these factors are included. Though it may be contentious to include these criteria in priority-setting, the results from this thesis indicate that the general public consider these criteria to be of importance.

10.3.3 Using respondents' demographic characteristics to predict preferences

OLS, SUR and the FML model were used to explore the relationship between the demographic characteristics of the random sample respondents and the criteria weights. Several relationships are found. For example, as might be expected, health care workers, respondents on low incomes and Maori place more importance on 'need' (relative to the other criteria) compared to respondents who do not work in health care, respondents on middle or high incomes and non-Maori. However, though there are several statistically significant relationships (at the 5% level) when the data are split in two the statistically significant results

are not common to both samples which suggests that the results may be by chance. It appears that the variation in respondents' preferences is largely idiosyncratic and not directly related to respondents' characteristics.

To explore the heterogeneity of respondents' preferences, the random sample respondents were clustered into groups based on the similarity of their criteria weights. Two different sets of clusters were explored. When respondents were clustered into two clusters, one cluster showed a much greater preference for 'need' compared to the other. When six clusters were formed, five of the clusters exhibited a definite preference for one or more of the criteria. For example, one cluster had a mean criterion weight for 'age' almost twice that of every other cluster. Some associations between the groups were found. However, instead of the 'main' demographic characteristics such as age, ethnicity, income or qualifications being associated with the clusters, health insurance, health usage, experience of a serious illness and work status were associated with the clusters. Most of the statistically significant effects were small and the associations generally involved a small number of respondents. Therefore, similar to the regression analysis, the demographic characteristics of respondents do not predict patterns of preferences.

10.4 Application of the prioritisation framework

The focus of this thesis is priority-setting at the meso-level of health care (i.e. the allocation of resources across a range of health and disability services). In New Zealand, the main agencies responsible for planning and prioritising publicly-funded health care at the meso-level are the MOH (guided by the NHC and other advisory committees), PHARMAC (who decides which medicines will be subsidised for use in public hospitals and in the community) and the 20 DHBs (who are responsible for purchasing and providing services for their geographical populations).

The NHC, as main advisor to the MOH, uses 11 criteria (discussed in Section 2.2.5) to support its decision-making. PHARMAC uses prioritisation criteria in conjunction with PBMA in deciding which medicines to fund. Both PHARMAC and the NHC recognise that as well as investment in new technologies, 'disinvestment' in some technologies is required if quality health care is to be delivered that is both affordable and sustainable (NHC 2012). Though an 'explicit disinvestment' approach is advocated by NHC (i.e. services are removed

or resources are shifted from low value areas to areas of higher value),[195] the NHC acknowledges that "the need for an explicit approach will depend on the sector's response to NHC advice" (p 7). DHBs are required to carry out principles-based prioritisation processes when allocating budgets[196] and are also required to meet the MOH's health targets.

The framework developed in this thesis can be used in a variety of ways: to allocate fixed budgets across a range of services, to re-allocate resources if the budget has been cut, to keep track of previous decisions especially when new information becomes available or to simply be used as a communication tool. The framework could be applied by the NHC, PHARMAC and DHBs in their funding decisions. However, as mentioned in Section 2.4.2, 'transparency' may not be welcomed by PHARMAC as some of their purchasing decisions involve combination purchasing, reduced prices and/or confidential rebates. In addition, PHARMAC deals with some very difficult funding decisions (usually involving high cost drugs for orphan diseases) and therefore it is doubtful that they would want the details made public, particularly in light of previous media coverage.

On the other hand, DHBs are publicly accountable to the regional populations they serve and therefore inclusion of key stakeholders in a prioritisation process is beneficial to all concerned. In terms of the proposed framework members of the community could be involved in each stage of the prioritisation process. For example, members of the general public, patients, health experts, private and public health care workers, and health managers could be involved in developing and determining the appropriate criteria and associated weights for prioritising services; health treatments could be categorised by experts in the area; and health managers and finance experts could allocate the budget across a range of health and disability services taking into account the criteria weights and any additional considerations.

Prioritising health services is always contentious as not everyone will be happy with the decisions made, particularly when services are reduced or completely removed. It is therefore important that the process is transparent and that care is taken at each stage of the process. For instance, as discussed in Section 6.3, the way health treatments are categorised will have a direct impact on the overall ranking of health treatments. Ideally there will be agreement

[195] According to the NHC, disinvestment rarely implies that a service or technology is completely removed. More often, disinvestment involves moving resources to interventions that are more cost-effective.
[196] DHBs are funded using the 'population based funding formula (PBFF)'. The amount each DBH receives is based on its share of the population, weighted by age, gender, ethnicity and deprivation, and is further adjusted for rural service delivery, overseas patients and unmet need.

among health experts regarding the categorisation of the treatments but as this step is fairly difficult (and based on the 'average' patient) this may not always be the case. However, an explicit prioritisation process ensures consistency and transparency, and when key stakeholders are involved throughout the process, there is likely to be a better understanding and acceptance of the decisions made.

The proposed prioritisation framework could also be adapted for use in other policy settings – for example, in specific treatment areas. (Point systems have already been developed in New Zealand (using 1000Minds and the PAPRIKA scoring method) in many treatment areas including coronary artery bypass graft surgery, hip and knee replacements and cataract surgery with further points systems planned for the future (Hansen et al 2012).) Indeed, the framework could potentially be used in other countries (at a local, regional or national level). The criteria, levels and weights are likely to be different but the methodological approach is generalisable (Baltussen 2006).

10.5 Limitations of the thesis

Priority-setting at the meso-level is the overarching focus of this thesis and as such participants in the focus groups were asked to rank 14 vignettes described at a treatment level. However, many participants commented, that to make it easier to rank the vignettes, they focussed on the number of patients being treated and ignored the other potential criteria. To avoid the same thing happening with the decision survey, 'number of patients' was considered as an 'additional consideration' and respondents were instead asked to choose between two *patients*. However, this blurs two levels of priority setting – meso and micro – and implies that the criteria weights obtained at the micro-level (i.e. choosing between two patients) can be used to prioritise treatments at the meso-level which may not be the case. For example, a respondent may not consider 'age' to be important when prioritising *within* a particular service, but they may consider it to be important when prioritising *across* services. This is a limitation of the research and something that could be considered for future research. For example, *treatments* could be included in the decision survey instead of *patients*.

As explained in Chapter 5, the criteria levels need to be ranked from lowest to highest for the decision survey. Most of the criteria levels are inherently ranked (e.g. 'small', 'medium' and 'large' for the criterion 'benefit to patient', with 'small' relating the lowest value and 'large' relating to the greatest value). For other criteria such as 'age' this is not the case. I ranked

'age' in ascending order of importance, i.e. 65+ years, 15-64 years and 0-14 years (reflecting the FIA). However, not everyone will agree with this ranking. Other software has a self-explication option (e.g. Sawtooth), although this raises other issues such as comparability of criteria weights.

Most of the respondents found the trade-off questions difficult to answer (66.5% of the snowball sample and 59.3% of the random sample). Choosing which patient to treat first is inherently difficult. As discussed in Section 5.7.1, though respondents were instructed to 'assume both patients are the same except as described', some respondents found it difficult not to think about additional criteria when answering the questions. Prioritising treatments and/or patients is complex and difficult. However, choosing between *two* patients described on just *two* criteria is much easier than choosing between *two or more* patients described on *four or more* criteria (which is typical of other DCE surveys).

At 10%, the response rate to the decision survey is low. Though the difference between respondents and non-respondents is considered to be small, the random sample is not completely representative and there is a potential non-response bias due to some respondents not having computer and/or internet access. Although there are definite advantages to using an online survey (as discussed in Chapter 5), it means that potential respondents are restricted to those with access to a computer and the internet, and/or who have confidence in completing an online survey.

10.6 Areas for future research

10.6.1 Compare methodologies

As discussed in Section 10.3.1, using 1000Minds and the PAPRIKA method to estimate respondents' preferences, is effective. Nevertheless, it would be interesting to compare how this method measures up against other methods such as AHP. For example, a random sample could be asked to rank 10 health treatments in order of priority. The sample would then be split in two with half of the sample completing a 1000Minds decision survey and the other half of the sample using an alternative method to generate the criteria weights. The resulting weights could then be compared to see which method produced weights that most closely resembled respondents' true preferences.

10.6.2 Repeat survey

The survey undertaken in this thesis could be a pilot for a larger nation-wide study with an emphasis on increasing the participation rate (i.e. reducing non-response bias) by contacting respondents directly for instance (email instead of letter), arranging access to the online survey if respondents do not have access to a computer and by offering an incentive (greater than $1) to do the survey.

In addition, as discussed in Section 10.3, most of the demographic characteristics that are associated with preferences are factors relating to health and work status rather than the main demographic characteristics of age and income. It is possible that the variation in preferences is due to factors not captured in this study. Repeating the survey with questions relating to respondents' health status, health care experience, place of employment and lifestyle characteristics such as smoking and exercise would provide additional information to further explore the association between respondents' lifestyle characteristics and the prioritisation criteria.

10.6.3 Age

In the decision survey the three levels for the criterion 'age' are ranked in ascending order of importance from old to young: 65+ years, 15-64 years and 0-14 years which broadly encompass three segments of society – retired people, working-aged people, and babies and children. Some respondents commented that the age groups are too wide. However, if the age groups are separated into smaller categories, what would be the order of preference? It would be interesting to investigate further New Zealanders' preferences in regard to age.

10.6.4 Lifestyle

Of the many comments received from the random sample respondents most related to the criterion 'lifestyle'. Given the depth of feeling relating to this criterion, it would be interesting to explore this area further.

10.7 Conclusion

The framework developed in this thesis illustrates how health care can be prioritised at the meso-level of health care funding in New Zealand. The framework brings together all the relevant information needed to prioritise services. It is up to the decision-makers to choose which treatments to fund within an available budget, but if decisions are made explicitly within a transparent framework that includes the preferences of key stakeholders, then there is likely to be greater public acceptance in the outcome. For a health system to be efficient, however, emphasis also needs to be placed on the delivery of, and access to, health services, otherwise people in need will still be left untreated.

~ References ~

Accent (2010) Review of Stated Preference and Willingness to Pay Methods. Report prepared by Accent in association with RAND Europe for the Competition Commission, London [online] Available at: www.competition-commission.org.uk/our_role/analysis/summary_and_report_combined.pdf.

Amaya-Amaya, M. Gerard, K. & Ryan, M. (2008) Discrete choice experiments in a nutshell. In: Using Discrete Choice Experiments to Value Health and Health Care. Ryan, M., Gerard, K. & Amaya-Amaya, M. (eds) Dordrecht, Springer, p 13-46.

AMGEN (2011) Novel Frameworks for Reimbursement Decision Making. [online] Available at: www.policymakers.ca/wordpress/wp-content/uploads/AMGEN-Novel-Frameworks-Jan-25th-2-pg.pdf.

Appleby, J., Devlin, N., Parkin, D., Buxton, M. & Chalkidou, K. (2009) Searching for cost effectiveness thresholds in the NHS. *Health Policy* 91(3): 239-245.

Arrow, K. J. (1963) Social Choice and Individual Values. (2nd ed) New York, John Wiley & Sons, Inc.

Ashton, T. (1999) The health reforms: to market and back? In: Redesigning the Welfare State: Problems, Policies and Prospects. Boston, J., Dalziel, P & St John S. (eds) Oxford University Press, p 134-153.

Ashton, T., Cumming, J. & Devlin, N. (2000) Priority-setting in New Zealand: translating principles into practice. *Journal of Health Services Research & Policy* 5(3): 170-175.

Badland, H. & Schofield, G. (2006) Test-Retest Reliability of a Survey to Measure Transport-Related Physical Activity in Adults. *Research Quarterly for Exercise and Sport.* 77(3): 386-390.

Bae, J., Joung, H., Kim, J.-Y., Nam K. K., Kim, Y.T. & Park, S-W. (2010) Test-Retest Reliability of a Questionnaire for the Korea Youth Risk Behavior Web-Based Survey. *Journal of Preventive Medicine and Public Health* 43(5):403-410.

Baltussen, R. & Niessen, L. (2006) Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation* 4(1): 14.

Baltussen, R., Stolk, E., Chisholm, D. & Aikens, M. (2006) Towards a multi-criteria approach for priority setting: an application to Ghana. *Health Economics* 15(7): 689-696.

Baltussen, R., ten Asbroek, A. H. A., Koolman, Shrestha, N., Bhattarai, p. & Niessen, L.W. (2007) Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? *Health Policy and Planning* 1-8.

Barbour, R. S. (1999) The Use of Focus Groups to Define Patient Needs. *Journal of Pediatric Gastroenterology and Nutrition* 28(4): S19-S22.

Barbour, R. (2005) Making sense of focus groups. Medical education 39(7): 742.

Bech, M. & Kristensen, M.B. (2009) Differential response rates in postal and Web-based surveys among older respondents. *Survey Research Methods* (3)1: 1-6.

Bech, M., Kjaer, T. & Lauridsen, J. (2011) Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Economics* 20(3): 273-286.

Belton, V. & Goodwin, P. (1996) Remarks on the application of the analytic hierarchy process to judgmental forecasting. *International Journal of Forecasting* (1): 155-161.

Belton, V. & Stewart, T. (2002) Multiple criteria decision analysis: an integrated approach. USA, Kluwer Academic Publishers.

Bender, D. & Ewbank, D. (1994) The focus group as a tool for health research: issues in design and analysis. *Health Transition Review* 4(1): 63-79.

Berney, L., Kelly, M., Doyal, L, Feder, G., Griffiths, C. & Jones, I.R. (2005) Ethical principles and the rationing of health care: a qualitative study in general practice. *British Journal of General Practice* 55: 620-625.

Biernacki, P. & Waldorf, D. (1981) Snowball Sampling, Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research* 10(2): 141-163.

Blasius, J. & Brandt, M. (2010) Representativeness in Online Surveys through Stratified Samples. *Bulletin de Methodologie Sociologique* 107(5): 106-108.

Bloomfield, A. (2003) New Zealand. In: Reasonable Rationing International Experience of Priority Setting in Health Care. Ham, C. & Robert, G. (eds) Philadelphia, Open University Press, p 16-41.

Bolloju, N. (2001) Aggregation of analytic hierarchy process models based on similarities in decision makers' preferences. *European Jouranl of Operational Research* 128(3): 499-508.

Boote, J., Baird, W. & Beecroft, C. (2010) Public involvement at the design stage of primary health research: A narrative review of case examples. Health Policy 95(1): 10-23.

Boulier, B. L., Datta, T. S. & Goldfarb, R.S. (2007) Vaccination Externalities. *The BE Journal of Economic Analysis and Policy* 7: Article 23.

Bowie, C., Richardson, A. & Skyes, W. (1995) Consulting the public about health service priorities. *BMJ* 311(7013): 1155.

Bowling, A. (1996) Health care rationing: the public's debate. *BMJ* 312(7032): 670-674.

Braveman, P. & Gruskin, S. (2003) Defining equity in health. *Journal of Epidemiology and Community Health* 57(4): 254-258.

Braveman, P. (2006) Health disparities and health equity: Concepts and Measurement. *Annual Review of Public Health* 27(1): 167-194.

Brawner, C.E., Felder, R.M., Allen, R.H., Brent, R. & Miller, T.K. (2001) A comparison of electronic surveying by e-mail and web. *Proceedings of the 2001 Annual ASEE Conference, ASEE.*

Bridges, F.P. (2003) Stated preference methods in health care evaluation: an emerging methodological paradigm in health economics. *Applied Health Economics and Health Policy* 2(4): 213-224.

Bridges, J. F. P., Hauber, A. B., Marshall, D., Lloyd, A., Prosser, L.A., Regier, D.A., Johnson, F.R. & Mauskopf, J. (2011) Conjoint Analysis Applications in Health—a Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health* 14(4): 403-413.

Brondani, M. A., MacEntee, M.I., Bryant, S.R. & O'Neill, B (2008) Using Written Vignettes in Focus Groups Among Older Adults to Discuss Oral Health as a Sensitive Topic. *Qualitative Health Research* 18(8): 1145-1153.

Brugha, C. (1998) Structuring and Weighting Criteria in Multi Criteria Decision Making (MCDM). In: Trends in Multicriteria Decision Making: Proceedings of the 13th International conference on Multiple Criteria Decision Making. Stewart, T.J. & Van den Honert, R.C. (eds) Springer-Verlag, p 229-242.

Bruni, R. A., Laupacis, A. & Martin, D. (2008) Public engagement in setting priorities in health care. *Canadian Medical Association Journal* 179(1): 15-18.

Bryan, S., Roberts, T., Heginbotham, C. & McCallum, A. (2002) QALY-maximisation and public preferences: results from a general population survey. *Health Economics* 11(8): 679-693.

Bryan, S. & Dolan, P. (2004) Discrete choice experiments in health economics. For better or for worse? *The European Journal of Health Economics* 5(3): 199-202.

Bryan, S. & Roberts T. (2008) Preferences for health care programmes. Results from a general population discrete choice survey. In: Using Discrete Choice Experiments to Value Health and Health Care. Ryan, M., Gerard, K. & Amaya-Amaya, M. (eds) Dordrecht, Springer, p 139-152.

Buckshaw, D. (2010) Decision Analysis Software Survey. [online] Available at: www.informs.org/ORMS-Today/Public-Articles/October-Volume-37-Number-5/Decision-Analysis-Software-Survey.

Buis, M.L. (2010) Relationship between the likelihood functions used in fmlogit and mlogit. [online] Available at: www.maartenbuis.nl/software/likelihoodFmlogit.pdf

Calltorp, J. (1999) Priority setting in health policy in Sweden and a comparison with Norway. *Health Policy* 50:1-22.

Cameron, T. & DeShazo, J. (2011) Differential attention to attributes in utility--theoretic choice models. *Journal of Choice Modelling* 3(3): 73-115.

Carey, M. A. (1995) Comment: Concerns in the Analysis of Focus Group Data. *Qualitative Health Research* 5(4): 487-495.

Carlsson, P. (2010) Priority setting in health care: Swedish efforts and experiences. *Scandinavian Journal of Public Health* 38(6): 561-564.

Casey, M.A. & Krueger, R. (2004) An overview of focus group interviewing. In: Evidence-based Practice Manual. Research and outcome measures in health and human sciences. Roberts, A. R. & Yeager, K. (eds) New York, Oxford University Press, p 61-70.

Charron, S. & Koechlin, E. (2010) Divided Representation of Concurrent Goals in the Human Frontal Lobes. *Science* 328(5976): 360-363.

Cherchi, E. & de Dios Ortuzar, J. (2006) Use of Mixed Revealed-Preference and Stated-Preference Models with Nonlinear Effects in Forecasting. *Transportation Research Record: Journal of the Transportation Research Board* (1977): 27-34

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences* (2nd ed) Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. T., Neumann, P. J. & Weinstein, M.C. (2008) Does Preventive Care Save Money? Health Economics and the Presidential Candidates. *New England Journal of Medicine* 358(7): 661-663.

Cook, C., Heath, F. & Thompson, R.L. (2000) A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement* 60(6): 821-836.

Cookson, R. & Dolan, P. (1999) Public views on health care rationing: a group discussion study. *Health Policy* 49(1-2): 63-74.

Cookson, R. & Dolan, P. (2000) Principles of justice in health care rationing. *Journal of Medical Ethics* 26(5): 323-329.

Cookson R. (2005) Evidence based policy making in health care: what it is and what is isn't. *Journal of Health Services Research and Policy* 10: 118-21.

Cookson, R., McCabe, C. & Tsuchiya, A. (2008) Public healthcare resource allocation and the Rule of Rescue. *Journal of Medical Ethics* 34(7): 540-544.

Couper, M.P., Kapteyn, A. &Schonlau, M. (2007) Noncoverage and Nonresponse in an Internet Survey. *Social Science Research* 36: 131–48.

Crawshaw, R., Garland, M., Hines, B & Anderson, B. (1990) Developing principles for prudent health care allocation. The continuing Oregon experiment. *Western Journal of Medicine* 152(4): 441.

Culyer, A. J. & Wagstaff, A. (1993) Equity and equality in health and health care. *Journal of Health Economics* 12(4): 431-457.

Culyer, A. (2001) Equity - some theory and its policy implications. *Journal of Medical Ethics* 27: 275-283.

Cumming, J. (1994) Core services and priority setting: the New Zealand experience. *Health Policy* 29(1&2): 41-60.

Daniels, N. & Sabin, J. (1998) The ethics of accountability in managed care reform. *Health Affairs* 17(5): 50-64.

Daniels, N. & Sabin, J. E. (2008) Accountability for reasonableness: an update. *BMJ* 337: a1850.

Dillman, D. (2007) Update. Mail and internet surveys: the tailored design method. (2nd ed) New Jersey, John Wiley & Sons, Inc.

Deaton, A. (2003) Health, Income and Inequality. *The National Bureau of Economic Research.* [online] Available at: http://www.nber.org/reporter/spring03/health.html

de Bekker-Grob, E. W., Ryan, M. & Gerard, K. (2012) Discrete choice experiments in health economics: a review of the literature. *Health Economics*: 21(2): 145-172. Epub 2010 Dec 19.

de Leeuw, E. & de Heer, W. (2002) Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. New York, John Wiley & Sons, Inc.

Denis, A., Simoens, S., Fostier, C., Mergaert, L. & Cleemput, I. (2009) Policies for Rare Diseases and Orphan Drugs. Health Technology Assessment. Brussels: Belgian Health Care Knowledge Centre, KCE reports 112C (D/2009/10.273/32)

Denzin, N. & Lincoln, Y. (1994) Handbook of qualitative research. London, Sage Publications Inc.

Detsky, A. S. & Laupacis, A. (2007) Relevance of Cost-effectiveness Analysis to Clinicians and Policy Makers. *JAMA* 298(2): 221-224.

Devlin, N., Hansen, P. & Herbison, P.(2000) Variations in self-reported health status: results from a New Zealand survey. *The New Zealand Medical Journal* 113(1123): 517-520.

Devlin, N. & Sussex, J. (2011) Incorporating multiple criteria in HTA: Methods and processes. London, Office of Health Economics.

Diederich, A., Winkelhage, J. & Wirsik, N. (2011) Age as a Criterion for Setting Priorities in Health Care? A Survey of the German Public View. *PLoS ONE* 6(8): e23930.

Dodgson, J., Spackman, M., Pearman, A. & Phillips, L. (2009) Multi-criteria analysis: a manual. London, UK., Department for Communities and Local Government.

Dolan, P. (1998) The measurement of individual utility and social welfare. *Journal of Health Economics* 17: 39-52.

Dolan, P., Cookson, R. & Ferguson, B. (1999) Effect of discussion and deliberation on the public's views of priority setting in health care: focus group study. *BMJ* 318(7188): 916-919.

Dolan, P., Shaw, R., Tsuchiya, A. & Williams, A. (2005) QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics* 14(2): 197-208.

Drummond, M., Stoddart, G., Torrance, G.W. & O'Brien, B.J. (2005) Methods for the Economic Evaluation of Health Care Programmes. New York, US, Oxford University Press.

Edwards, P., Roberts, I., Clarke, M.J., DiGuiseppi, C., Wentz, R., Kwan, I., Cooper, R., Felix, L.M. & Pratap, S. (2009) Methods to increase response to postal and electronic questionnaires. Cochrane Database of Systematic Reviews (3)

Edwards, W. (1977) How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC-7: 326-340.

Edwards, W. & Barron, F.H. (1994) SMARTS and SMARTER: Improved simple methods for multiattribute utility measurements. *Organizational Behavior and Human Decision Processes* 60: 306-325.

Farrant, A. (2009) The fair innings argument and increasing life spans. *Journal of Medical Ethics* 35(1): 53-56.

Ferreira, J. E. d-S. & Veiga, D.V.d. (2008) Test-retest reliability of a simplified questionnaire for screening adolescents with risk behaviours for eating disorders in epidemiologic studies. *Brazilian Journal of Epidemiology* 11(3): 1-9.

Field, A. (2009) Discovering Statistics Using SPSS. (3rd ed) Sage Publications Inc.

Figueira, J., Greco, S., & Ehrgott, M. (eds) (2005) Multiple Criteria Decision Analysis: State of the Art Surveys. London, Springer Verlag.

Finch, H., & Lewis, J. (2003) Focus Groups. In: *Qualitative research practice: A guide for social science students and researchers.* Ritchie, J. & Lewis, J. (eds) London, Sage Publications Ltd, p 170-198.

Fishbein, M. A. (1963) An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human Relations*16, 233-239.

Florin, D. & Dixon, J. (2004) Public involvement in health care. *BMJ* 328(7432): 159-161.

Flynn, T. N., Louviere, J. J., Peters, T.J. & Coast, J. (2007) Best-worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics* 26(1): 171-189.

Forman, E. & Selly, M. (2001) Decision by objectives: how to convince others that you are right. Singapore, World Scientific.

Fülöp, J. (2005) Introduction to Decision Making Methods. Laboratory of Operations Research and Decision Systems: Computer and Automation Institute, Hungarian Academy of Sciences 1.

Gafni, A., Birch, S. & Mehrez, A. (1993) Economics, health and health economics: HYEs versus QALYs. *Journal of Health Economics* 12: 325-339.

Gallego, G., Hass, M., Hall, J. & Viney, R. (2010) Reducing the use of ineffective health care interventions. A report by the Centre for Health Economics Research and Evaluation for NSW Treasury. Sydney, University of Technology. Working Paper 2010/5.

Gauld, R. & Derrett. S. (2000) Solving the surgical waiting list problem? New Zealand's booking system. *The International Journal of Health Planning and Management* 15(4): 259-272.

Gerard, K., Ryan, M. & Amaya-Amaya, M. (2008) Introduction. In: Using Discrete Choice Experiments to Value Health and Health Care. Ryan, M., Gerard, K. & Amaya-Amaya, M. (eds) Dordrecht, Springer, p 1-10.

Glaser, B. G., Strauss, A. L. & Strutzel, E. (1967) The Discovery of Grounded Theory; Strategies for Qualitative Research. *Nursing Research* 17(4): 364.

Golan, O. & Hansen, P. (2008) A new framework to help Israel's Health Basket Committee choose new health technologies. The VfM (Value for Money) Chart. Note, www.gertnerinst.org.il/health.

Golan, O., Hansen, P., Kaplan, G. & Tal, O. (2010) Health technology prioritization: Which criteria for prioritizing new technologies and what are their relative weights? *Health Policy* 102(2-3): 126-135.

Golan, O. & Hansen, P. (forthcoming) Which health technologies should be funded? A prioritization framework based explicitly on value for money. *Israel Journal of Health Policy Research*.

GRADE working group. (2004) Grading quality of evidence and strength of recommendations. *BMJ* 328(7454): 1490.

Greene, W.H. (2000) Econometric Analysis. 4th ed. New Jersey, Prentice-Hall Inc.

Grocott, R. (2009) Applying Programme Budgeting Marginal Analysis in the health sector: 12 years of experience. *Expert Review of Pharmacoeconomics and Outcomes Research* 9(2): 181-187.

Guba, E. (1981) Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Technology Research and Development* 29(2): 75-91.

Guyatt, G. H., Oxman, A. D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coelllo, P. & Schunemann, H.J. (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336(7650): 924-926.

Hadorn, D. (2000) Setting priorities for waiting lists: defining our terms. *Canadian Medical Association Journal* 163(7): 857-860.

Haffner, M. E. (2006) Adopting Orphan Drugs -- Two Dozen Years of Treating Rare Diseases. *New England Journal of Medicine* 354(5): 445-447.

Hair, J.F., Black, W.C. & Babin, B.J. (2010) Multivariate Data Analysis: A Global Perspective. (7th ed) Upper Saddle River, NJ, London, Pearson Education Inc.

Ham, C. (1997) Priority setting in health care: learning from international experience. *Health Policy* 42(1): 49-66.

Hansen, P. (2002) Another Graphical Proof of Arrow's Impossibility Theorem. *Summer* 217:235.

Hansen P. (2006) A Theoretical review of PHARMAC's over-arching approach to deciding which pharmaceuticals to fund, including high cost ones. Department Economics, University Otago, Dunedin, New Zealand.

Hansen, P. & Ombler, F. (2008) A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis* 15(3-4): 87-107.

Hansen, P., Hendry, A., Naden, R., Ombler, F. & Stewart R. (2012) A new process for creating points systems for prioritizing patients for elective health services. *Clinical Governance: An International Journal* 17(3): 200-209.

Harris, J. (1985) The value of life. London, Routledge.

Harris, R., Tobias, M., Jeffreys, M., Waldegrave, K., Karlsen, S. & Nazroo, J. (2006) Effects of self-reported racial discrimination and deprivation on Maori health and inequalities in New Zealand: cross-sectional study. *The Lancet* 367(9527): 2005-2009.

Hastie, R. & Dawes, R.M. (2010) Rational Choice in an Uncertain World. The Psychology of Judgment and Decision Making. (2nd ed) USA, Sage Publications Inc.

Hauck, K., Goddard, M. & Smith, P.C. (2004) The economics of priority setting for health care: a literature review. International Bank for Reconstruction and Development, World Bank.

Hayes, B.E. (2008) Measuring Customer Satisfaction and Loyalty: Survey Design, Use, and Statistical Analysis Methods. (3rd ed) Milwaukee, WI, ASQ, Quality Press.

Healey, B., Macpherson, T. & Kuijten, B. (2005) An Empirical Evaluation of Three Web Survey Design Principles. *Marketing Bulletin* 16 (Research Note 2): 1-9.

Health Funding Authority (2000) Overview of the Health Funding Authority's Prioritisation Decision Making Framework. Health Funding Authority, Wellington.

Health Funds Association of New Zealand (2010) Fact File – Health Insurance in New Zealand. [online] Available at: www.healthfunds.org.nz/pdf/Fact%20File%20oct2010 .pdf

Hughes, D. A., Tunnage, B. & Yeo, S.T. (2005) Drugs for exceptionally rare diseases: do they deserve special status for funding? *Quarterly Journal of Medicine* 98(11): 829-836.

Hughes, R. & Huby, M. (2002) The application of vignettes in social and nursing research. *Journal of Advanced Nursing* 37(4): 382-386.

James, C., Carrin, G., Savedoff, W. & Hanvoravongchai, P. (2005) Clarifying Efficiency-Equity Tradeoffs Through Explicit Criteria, With a Focus on Developing Countries. *Health Care Analysis* 13(1): 33-51.

Jan, S., Mooney, G., Ryan, M., Bruggemann, K. & Alexander, K. (2000) The use of conjoint analysis to elicit community preferences in public health research: a case study of hospital services in South Australia. *Australian and New Zealand Journal of Public Health* 24(1): 64-70.

Johnson, R.M.. (2001) History of ACA. Sawtooth Software Inc. Research paper series.

Joint DHB & Ministry of Health Working Group on Prioritisation (2005) The Best Use of Available Resources. An approach to prioritization. Ministry of Health, Wellington.

Jordan, J., Dowswell, T., Harrison, S., Lilford, R.J. & Mort, M. (1998) Health needs assessment: Whose priorities? Listening to users and the public. *BMJ* 316(7145): 1668-1670.

Jonsson, E. & Banta, D. (1999) Management of health technologies: an international view. *BMJ* 319: 1293-1295.

Jusot, F., Tubeuf, S. & Trannoy, A. (2010) Effort or circumstances: which one matters in health inequality? IRDES, Paris: Working Paper DT33.

Kapiriri, L. & Frithjof Norheim, O. (2004) Criteria for priority-setting in health care in Uganda: exploration of stakeholders' values. *Bulletin of the World Health Organisation* 82(3): 172-179.

Kenny, N. & Joffres, C. (2008) An Ethical Analysis of International Health Priority-Setting. *Health Care Analysis* 16(2): 145-160.

Keeney, R.L. & Raiffa, H. (1976) Decisions with Multiple Objectives: Performances and Value Trade-Offs. New York, JohnWiley & Sons, Inc.

Kidd, P. S. &. Parshall, M. B (2000) Getting the Focus and the Group: Enhancing Analytical Rigor in Focus Group Research. *Qualitative Health Research* 10(3): 293-308.

Kingi, T. (2007) The Treaty of Waitangi: A framework for Maori health development. *New Zealand Journal of Occupational Therapy* 54(1): 4-10.

Kitzinger, J. (1994) The methodology of focus groups: the importance of interaction between research participants. *Sociology of health and illness* 16(1): 103-121.

Kitzinger, J. (1995) Introducing Focus Groups. *BMJ* 311(7000): 299-302.

Koch, S.F. (2010) Fractional Multinomial Response Models with an Application to Expenditure Shares. University of Pretoria, Department of Economics Working Paper Series 2010-21.

Krantz, D. H. (1972) Measurement Structures and Psychological Laws. *Science* 175(4029): 1427-1435.

Krippendorff, K. (2004) Content Analysis, An Introduction to its Methodology. USA, Sage Publications, Inc.

Krueger, R.A. & Casey, M.A. (2000) Focus groups: A practical guide for applied research. (3rd ed) Thousand Oaks, CA, Sage Publications Inc.

Kvale, S. (1996) Inter Views: An introduction to qualitative research interviewing. Thousand Oaks, CA, Sage Publications Inc.

Kypri, K., Samaranayaka, A., Connor, J., Langley, J. D., & MacLennan, B. (2011) Non-response bias in a web-based health behaviour survey of New Zealand tertiary students. *Preventive Medicine*. Advance online publication. doi: 10.1016/j.ypmed.2011.07.017

Lancaster, K. J. (1966) A New Approach to Consumer Theory. *The Journal of Political Economy* 74(2): 132-157.

Lancsar, E. & Louviere, J. (2006) Deleting irrational responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Economics* 15(8): 797-811.

Landon, E. L., Jr. (1971) Order Bias, the Ideal Rating, and the Semantic Differential. *Journal of Marketing Research* 8(3): 375-378.

Lenaghan, J. (1999) Involving the public in rationing decisions. The experience of citizens juries. *Health Policy* 49(1-2): 45-61.

Lewins, A. & Silver, C. (2007) Using Software in Qualitative Research: A Step-by-Step Guide**.** Los Angeles, Sage Publications Inc.

Lincoln, Y. &. Guba, E (1985) Naturalistic Inquiry. USA, Sage Publications Inc.

Litva, A., Coast, J., Donovan, J., Eyles, J., Shepherd, M., Tacchi, J., Abelson, J. & Morgan, K. (2002) The public is too subjective: public involvement at different levels of health-care decision making. *Social Science & Medicine* 54(12): 1825-1837.

Logan, R., Fougere, G., Hague, K., Haretuku, R., Holloway, L., Moore, A., Page-Carruth, A., Pearce, N., Stewart, L., Sullivan, L., Talemaitoga, A. & Tepania-Palmer, G. (2004) Prioritising Health Services. A background paper for the National Health Committee, National Health Committee.

Louviere, J., Flynne, T. & Carson, R.T. (2010) Discrete Choice Experiments Are Not Conjoint Analysis. *Journal of Choice Modelling* 3(3): 57-72.

Luce, D.R. (1992) A Theory of Certainty Equivalents for Uncertain Alternatives. *Journal of Behavioral Decision Making* 5: 201-216.

Manski, C.F. (1977) The structure of random utility models. *Theory and Decision* 8:229-254.

Manski, C. F. (1999) Analysis of Choice Expectations in Incomplete Scenarios. *Journal of Risk and Uncertainty* 19(1): 49-66.

Martire, L.M., Lustig, A.P., Schulz, R., Miller, G.E. & Helgeson, V.S. (2004) Is it beneficial to involve a family member? A meta-analysis of psychosocial interventions for chronic illness. *Health Psychology* 23(6): 599-611.

Mays, N. & Pope, C. (2000) Qualitative research in health care: Assessing quality in qualitative research. *BMJ* 320(7226): 50-52.

McCabe, C., Claxton, K. & Tsuchiya, A. (2005) Orphan drugs and the NHS: should we value rarity? *BMJ* 331(7523): 1016-1019.

McCormack, P., Quigley, J. & Hansen, P. (2010) Review of Access to High-cost, Highly-Specialised Medicines in New Zealand, Report to Minister of Health, Hon Tony Ryall.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: Frontiers in Econometrics. Zarembka, P. (ed) New York: Academic Press, p 105-142.

McFadden, D., Bemmaor, A., Caro, F., Dominitz, J., Byung-Hill, J., Lewbel, A., Matzkin, R., Molinari, F., Schwarz, N., Willis, R. & Winter, J. (2005) Statistical Analysis of Choice Experiments and Surveys. *Marketing Letters* 16(3): 183-196.

Menon, D., Stafinski, T. & Douglas, M. (2007) Priority-setting for healthcare: who, how, and is it fair? *Health Policy* 84(2-3): 220-233.

Miller, E.T., Neal, D.J., Roberts, L.J., Baer, J.S., Cressler, S.O., Metrik, J. & Marlatt, G.A. (2002) Test-retest reliability of alcohol measures: is there a difference between internet-based assessment and traditional methods? *Psychology of Addictive Behaviors* 16(1): 56-63.

Millet, I. & Saaty, T. L. (2000) On the relativity of relative measures – accommodating both rank preservation and rank reversals in the AHP. *European Journal of Operational Research* 121(1): 205-212.

Minister of Health (2000) The New Zealand Health Strategy. Ministry of Health, Wellington.

Mitton, C. R. & Donaldson, C. (2003) Setting priorities and allocating resources in health regions: lessons from a project evaluating program budgeting and marginal analysis (PBMA) *Health Policy* 64(3): 335-348.

Mitton, C. & Patten, S. (2004) Evidence-based priority-setting: what do the decision-makers think? *Journal of Health Services Research Policy* 9(3): 146-152.

Mitton, C., Smith, N., Peacock, S., Evoy, B. & Abelson, J. (2009) Public participation in health care priority setting: A scoping review. *Health Policy 91(3): 219-228.*

Mooi, E. & Sarstedt, M. (2011) A Concise Guide to Market Research. Berlin Heidelberg, Springer-Verlag.

Morse, J. M. (1995) The Significance of Saturation. *Qualitative Health Research* 5(2): 147-149.

Morse, J. M., Barrett, M., Mayan, M., Olson, K & Spiers, J. (2002) Verification Strategies for Establishing Reliability and Validity in Qualitative Research. *International Journal of Qualitative Methods* 1**:** 1-19.

Moshkovich, H. M., Mechitov, A. I. & Olson, D.L.. (2002) Ordinal judgments in multiattribute decision analysis. *European Journal of Operational Research* 137(3): 625-641.

Mullahy, J. & Robert, S.A. (2010) No time to lose? Time constraints and physical activity in the production of health. *Review of Economics of the Household* 8(4): 409-432.

Mullen, P. (2004) Quantifying priorities in healthcare: transparency or illusion? *Health Services Management Research* 17(1): 47.

Nair, K., Dolovich, L., Cassels, A., McCormack, J., Levine, M., Gray, J., Mann, K. & Burns, S. (2002) What patients want to know about their medications. Focus group study of patient and clinician perspectives. *Canadian Family Physician* 48(1): 104-110.

National Health Committee (2012) Think Piece: Towards a plan for better targeting of existing health services. National Health Committee, Wellington.

National Institute for Clinical Excellence. NHS should consider paying premium prices for drugs to treat patients with very rare diseases says NICE Citizen's Council. Press release 26 Jan. www.nice.org.uk/pdf/2005_002_CitizensCouncilreport.pdf.

Neogi, T., Aletaha, D. et al. (39 authors) (2010) The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: Phase 2 methodological report. *Arthritis & Rheumatism* 62(9): 2582-2591.

Neumann, P.J., Goldie, S.J. & Weinstein, M.C. (2000) Preference-Based Measures in Economic Evaluation in Health Care. *Annual Review of Public Health* 21: 587-611.

Neumann, P.J. (2009) Lessons for Health Technology Assessment: It is not only about the evidence. *International Society for Pharmacoeconomics and Outcomes Research* 12(2): S45-S48.

New Zealand Treasury (March 2005) Population Ageing and Government Health Expenditure. New Zealand Treasury Policy Perspectives Paper 05/01.

New Zealand Treasury (2011) A Snapshot of the 2011 Financial Statements of the New Zealand Government. Published 11 October 2011.

Nord, E., Richardson, J., Street, A., Kuhse, H. & Singer, P. (1995) Maximizing health benefits vs egalitarianism: An Australian survey of health issues. *Social Science & Medicine* 41(10): 1429-1437.

Nord, E., Street, A., Richardson, J., Kuhse, H. & Singer, P. (1996) The significance of age and duration of effect in social evaluation of health care. *Health Care Analysis* 4(2): 103-111.

Nord, E. (2005) Concerns for the worse off: fair innings versus severity. *Social Science & Medicine* 60(2): 257-263.

Norusis, M. (2009) PASW Statistics 18 Statistical Procedures Companion. Upper Saddle River, NJ, Prentice Hall Inc.

Noseworthy, T. W., McGurran, J. J. & Hadorn, D.C. (2003) Waiting for scheduled services in Canada: development of priority-setting scoring systems. *Journal of Evaluation in Clinical Practice* 9(1): 23-31.

Nulty, D.D. (2008) The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education* 33(3): 301-314.

O'Donnell, J. C., Pham, S. V., Pashos, C.L., Miller, D.W. & Smith, M.D. (2009) Health Technology Assessment: Lessons Learned from Around the World - An Overview. Value in Health 12: S1-S5.

Olsen, J. A. (1997) Aiding priority setting in health care: is there a role for the contingent valuation method? *Health Economics* 6(6): 603-612.

Olson, C.L. (1979) Practical considerations in choosing a MANOVA test statistic: A rejoinder to Stevens. *Psychological Bulletin* 86: 1350-1352.

Olson, K. & Spiers, J. (2002) Verification Strategies for Establishing Reliability and Validity in Qualitative Research. *International Journal of Qualitative Met*hods 1**:** 1-19.

Ombler, F. & Hansen, P. 1000Minds software, 2012. Available from www.1000minds.com.

Papke, L. E. & Wooldridge, J.M. (1996) Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics* 11(6): 619-632.

Peacock, S., Ruta, D., Mitton, C., Donaldson, C., Bate, A. & Murtagh, M. (2006) Using economics to set pragmatic and ethical priorities. *BMJ* 322:482-485.

Peacock, S., Mitton, C., Bate, A., McCoy, B. & Donaldson, C. (2009) Overcoming barriers to priority setting using interdisciplinary methods. *Health Policy* 92(2–3): 124-132.

Perkins, M. R. V., Devlin, N. J. & Hansen P. (2004) The validity and reliability of EQ-5D health state valuations in a survey of Māori. *Quality of Life Research* 13(1): 271-274.

Perreault Jr, W. D. (1975) Controlling order-effect bias. *Public Opinion Quarterly* 39(4): 544.

Persad, G., Wertheimer, A. & Emanuel, E.J. (2009) Principles for allocation of scarce medical interventions. *Lancet* 373: 423-431.

Pickler, R. H. Evaluating Qualitative Research Studies. *Journal of Pediatric Health Care* 21(3): 195-197.

Rabiee, F. (2007) Focus-group interview and data analysis. *Proceedings of the nutrition society* 63(04): 655-660.

Radhakrishna, R.B. (2007) Tips for Developing and Testing Questionnaires/Instruments. *Journal of Extension* [Online] 45(1). Available at: www.joe.org/joe/2007february /tt2.php.

Raftery, J. (2001) NICE: faster access to modern treatments? Analysis of guidance on health technologies. *BMJ* 323(7324): 1300-1303.

Ramsey, J.B. (1969) Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society, Series B* 31(2), 350–371.

Rawlins, M., Barnett, D. & Stevens, A. (2010) Pharmacoeconomics: NICE's approach to decision-making. *British Journal of Clinical Pharmacology* 70(3): 346-349.

Reilly, R.M. (2010) Monoclonal Antibody and Peptide-Targeted Radiotherapy of Cancer. (ed) New Jersey, John Wiley & Sons, Inc.

Rhodes, R., Battin, M. & Silvers, A. (2002) Medicine and social justice: essays on the distribution of health care. New York, Oxford University Press.

Ritchie, J. & Spencer, L. (1994) Qualitative data analysis for applied policy research. In: Analyzing qualitative data. Bryman, A. & Burgess, R. G. (eds) London, Routledge, p 173-194.

Rivlin, M. (2000) Why the fair innings argument is not persuasive. *BMC Medical Ethics* 1(1): 1.

Robson, B. & Harris, R. (eds) (2007) Hauora: Maori Standards of Health IV. A study of the years 2000-2005.Wellington, Te Ropu Rangahau Hauora a Eru Pomare.

Rosen, B. (2011) Healthcare in Israel for U.S. audiences. Companion report to Healthcare in the US and Israel: Comparative Overview. Myers-JDC-Brookdale Institute, Smokler Center for Health Policy Research.

Roy, B. (1968) Classement et choix en presence de points de vue multiples (la methode ELECTRE) *RIRO* 8, 57Ð75.

Russell, C. K. & Gregory, D. M. (2003) Evaluation of qualitative research studies. *Evidence Based Nursing* 6(2): 36-40.

Russell L.B. (2007) *Prevention's Potential for Slowing the Growth of Medical Spending*. Washington DC, National Coalition on Health Care.

Ruta, D., Mitton, C., Bate, A. & Donaldson, C. (2005). Programme budgeting and marginal analysis: bridging the divide between doctors and managers. *BMJ* **330**(7506): 1501-1503.

Ryan, M. (1999) A role for conjoint analysis in technology assessment in health care? *International Journal of Technology Assessment in Health Care* 15(03): 443-457.

Ryan, M. & Farrar, S. (2000) Using conjoint analysis to elicit preferences for health care. *BMJ* 320(7248): 1530-1533.

Ryan, M., Bate, A., Eastmond, C.J. & Ludbrook, A. (2001) Use of discrete choice experiments to elicit preferences. *Qualitative Health Care* 10(90001): i55-60.

Ryan, M. & Gerard, K. (2003) Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied Economics and Health Policy* 2(55-64)

Ryan, M. (2004) Discrete choice experiments in health care. *BMJ* 328(7436): 360-361.

Saaty, T. (1994) How to make a decision: the analytic hierarchy process. *Interfaces*: 19-43.

Saaty, T.L. (2008) Decision making with the analytic hierarchy process. *International Journal of Services Sciences* 1(1): 83-98.

Sabik, L. & Lie, R. (2008) Priority setting in health care: Lessons from the experiences of eight countries. *International Journal for Equity in Health* 7(1): 4.

Sampietro-Colom, L., Espallargues, M., Rodriguez, E., Comas, M., Alonso, J., Castells, X. & Pinto, J.L. (2008) Wide Social Participation in Prioritizing Patients on Waiting Lists for Joint Replacement: A Conjoint Analysis. *Medical Decision Making* 28(4): 554-566.

Samuelson, P.A. (1948) Consumption Theory in Terms of Revealed Preference. *Economica* 15(60): 243-253.

Schlademann, S., Meyer, T. & Raspe, H. (2008) The test-retest reliability of a questionnaire on the occurrence and severity of back pain in a German population sample. *International journal of public health* 53(2): 96-103.

Schoenberg, N. E. & Ravdal, H. (2000) Using vignettes in awareness and attitudinal research. International Journal of Social Research Methodology, Routledge. 3**:** 63-74.

Scott, A. (2002) Identifying and analysing dominant preferences in discrete choice experiments: An application in health care. *Journal of Economic Psychology* 23(3): 383-398.

Scott, M.J. & Antonsson, M.J. (2000) Arrow's Theorem and Engineering Design Decision Making. *Research in Engineering Design* 11(4): 218-228.

Shah, K. K. (2009) Severity of illness and priority setting in healthcare: A review of the literature. *Health Policy* 93(2-3): 77-84.

Shani, S., Siebzehner,M.I., Luxenburg, O. & Shemer, J. (2000) Setting priorities for the adoption of health technologies on a national level -- the Israeli experience. *Health Policy* 54(3): 169-185.

Sheerin I., Green T., Sellman D., Adamson S. & Deering D. (2004) Reduction in crime and imprisonment for injecting drug users on a Methadone Maintenance Programme in New Zealand. *The New Zealand Medical Journal* 117(1190).

Singer, P. A., Martin, D.K., Giacomini, M. & Purdy, C. (2000) Priority setting for new technologies in medicine: qualitative case study. *BMJ* 321(7272): 1316-1318.

Signal, L., Martin, J. Reid, P., Carroll, C., Howden-Chapman, P., Ormsby, V, Richards, R., Robson, B. & Wall, T. (2007) Tackling health inequalities: moving theory to action. *International Journal for Equity in Health* 6(1): 12.

Sivakumar, A. & Bhat, C.R. (2002) A fractional split distribution model for statewide commodity flow analysis. *Transportation Research Record* 1790: 80-88.

Sivo, S.A., Saunders, C., Chang, Q., & Jiang, J.J. (2006) How Low Should You Go? Low Response Rates and the Validity of Inference in IS Questionnaire Research. *Journal of the Association for Information Systems* 7(6): Article 17.

Slovic, P. (Aug 1975) Choice between equally valued alternatives. *Journal of Experimental Psychology: Human Perception and Performance* 1(3): 280-287.

Smee, A. & Brennan, M. (2000) Electronic surveys: A comparison of e-mail, web and mail. Proceedings from ANZMAC 2000, Visionary Marketing for the 21st Century Facing the Challenge. Griffith University, Australia.

Smith, K. & Fennessy, P. (2011) The use of conjoint analysis to determine the relative importance of specific traits as selection criteria for the improvement of perennial pasture species in Australia. *Crop & Pasture Science* 62: 355-365.

Smith, R. D. (2003) Construction of the contingent valuation market in health care: a critical assessment. *Health Economics* 12(8): 609-628.

Spalding, N. J. & Phillips, T. (2007) Exploring the Use of Vignettes: From Validity to Trustworthiness. *Qualitative Health Research* 17(7): 954-962.

Stafinski, T., Menon, D., Philippon, D. & McCabe, C. (2011) Health Technology Funding Decision-Making Processes Around the World. *PharmacoEconomics* 29(6): 475-495.

Statistics New Zealand (2009) Household Use of Information and Communication Technology: 2009. [online] Available at: www.stats.govt.nz/browse_for_stats/industry_sectors/information_technology_and_communications/HouseholdUseofICT_HOTP2009/Commentary.aspx

Statistics New Zealand (2011) National Accounts: Year ended March 2011 - Statistics New Zealand. [online] Available at: www.stats.govt.nz/~/media/Statistics/Browse%20for%20stats/NationalAccounts/HOTPYeMar11/NationalAccountsYeMar11HOTP.pdf

Statistics New Zealand (2011) Index of key New Zealand statistics. [online] Available at: www.stats.govt.nz/browse_for_stats/snapshots-of-nz/index-key-statistics/population.aspx

Stevens, S.S. (1946) On the Theory of Scales of Measurement. *Science* 103(2684):677-680.

Stewart, D., Shamdasani, P.N. & Rook, D.W. (2007) Focus Groups: Theory and Practice. (2nd ed) Sage Publications Ltd.

Stolk, E. A., Pickee, S. J., Ament, A.H.J.A. & Busschbach, J.J.V. (2005) Equity in health care prioritisation: An empirical inquiry into social value. *Health Policy* 74(3): 343-355.

Stolk, E.A. & Poley, M.J. (2005) Criteria for determining a basic health services package. Recent developments in The Netherlands. *European Journal of Health Economics* 50:2-7.

Stronks, K., Strijbis, A.-M., Wendte, J.F. & Gunning-Schepers, L.J. (1997) Who should decide? Qualitative analysis of panel data from public, patients, healthcare professionals, and insurers on priorities in health care. *BMJ* 315(7100): 92-96.

Swartz, K. (2009) Health care for the poor: For whom, what care, and whose responsibility? *Focus* 26(2): 69-74.

Tabachnick, B.G. & Fidell, L.S. (2007) Using Multivariate Statistics. (5th ed) USA, Pearson Education Inc.

Teng, F., Mitton, C. & MacKenzie, J. (2007) Priority setting in the provincial health services authority: survey of key decision makers. *BMC Health Services Research* 7(1): 84.

Towers, A. (2008) Health, work and retirement survey summary report for the 2006 data wave: Methodology. Available from Massey University, Health, Work, and Retirement Survey Website: http://hwr/resources/methodology_towers.pdf.

Treadwell, J., Tregear, S., Reston, J. & Turkelson, C. (2006) A system for rating the stability and strength of medical evidence. *BMC Medical Research Methodology* 6(1): 52.

Tversky, A., Sattath, S. & Slovic, P. (Jul 1988) Contingent weighting in judgment and choice. *Psychological Review* 95(3): 371-384.

van Manen, M. (1990) Researching Lived Experience: Human Science for an Action Sensitive Pedagogy. Ontario, Althouse Press.

Von Winterfeldt, D. & Edwards, W. (1986) Decision analysis and behavioral research. Cambridge, New York, Cambridge University Press.

Ye, X. & Pendyala, R.M. (2005) A model of daily time use allocation using the fractional logit methodology. In Transportation Research Board 2005 Annual Meeting CD-ROM.

Walley, T., Rawlins, M. & Stein, K. (1998) Health technology assessment – the role of the pharmaceutical panel. *British Journal of Clinical Pharmacology* 45: 217-220.

Warden, B.A & Wong, S. (2007) Introduction to Qualitative Data Analysis, Constella Group. Accessed Nov 2010 from www.eval.org/SummerInstitute07/Handouts.

Waternaux, C.M. (1976) Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika* 63(3): 639-645.

Wensing, M. & Elwyn, G. (2003) Improving the quality of health care: Methods for incorporating patients' views in health care. *BMJ* 326(7394): 877-879.

Whitehead, M. (1991) The concepts and principles of equity and health. *Health Promotion International* 6(3): 217-228.

Wilks, D.S. (1995) Statistical Methods in the Atmospheric Sciences. California, Academic Press.

Williams, A. (1997) Intergenerational Equity: An Exploration of the 'Fair Innings' Argument. *Health Economics* 6(2): 117-132.

Wilson, E., Rees, J. & Fordham, R. (2006) Developing a prioritisation framework in an English Primary Care Trust. *Cost Effectiveness and Resource Allocation* 4(1): 3.

Wiseman, V., Mooney, G., Berry, G. & Tang, K.C. (2003) Involving the general public in priority setting: experiences from Australia. *Social Science & Medicine* 56(5): 1001-1012.

Woodward, A. & Kawachi, I. (2000) Why reduce health inequalities? *Journal of Epidemiology and Community Health* 54(12): 923-929.