A Practical Approach to Well-being Based Policy Development: What Do New Zealanders Want from Their Retirement Income Policies?

Joey Au, Andrew Coleman and Trudy Sullivan*

Paper prepared for the 56th NZAE Conference, Wellington, 1 – 3 July 2015

DRAFT

SUBJECT TO REVISION AND NOT FOR QUOTATION WITHOUT PERMISSION OF THE AUTHORS

* Authors contact details:
joey.au@treasury.govt.nz
andrew.coleman@otago.ac.nz
trudy.sullivan@otago.ac.nz

Abstract

This paper investigates the practicality of using a sophisticated multi-criteria analysis technique to estimate the preferences of a representative sample of the public to inform policy advice. Our application concerns retirement income policy and we use a multi-criteria decision-making survey to (i) investigate the relative importance of seven aspects of retirement income policies to a sample of 1,066 New Zealanders, (ii) document the diversity of policy preferences in a statistically rigorous manner, and (iii) evaluate the way people rank three different retirement income policies from an individual well-being perspective. The results of the paper suggest that multi-criteria surveys as a tool have considerable potential to help policymakers develop and identify policies that are aligned with the way people want to live. In terms of retirement income policies, we find that (i) there is widespread opposition to means-testing, (ii) a majority of respondents would choose an increase in current taxes if this could prevent even larger tax increases on future generations, and (iii) there are strongly divergent preferences over the appropriate eligibility age for New Zealand Superannuation. Overall, a policy combination that raises the age of eligibility for New Zealand Superannuation and reduces future tax increases is opposed by many and preferred by few. However, a policy that more aggressively prefunds New Zealand Superannuation by immediately raising taxes is supported by a majority of people of all ages and income groups.

Executive Summary

"*The ultimate purpose of public policy is to enable people to pursue better lives - to enhance the capabilities and opportunities for people live the kind of lives they have reason to value.*"

The New Zealand Treasury's Living Standards Framework specifies that the primary purpose of public policy is to enhance the capabilities and opportunities of individuals to pursue the lives they have reason to value. To design public policy that enhances individual well-being we need to know which aspects of well-being are most important to people. The purpose of this paper is to develop and apply a method for undertaking such an assessment in the specific context of retirement income policy.

This paper investigates the practicality of using a sophisticated multi-criteria analysis technique to estimate the preferences of a representative sample of the public in a manner that can inform policy advice. In particular, it uses a multi-criteria decision-making survey to (i) investigate the relative importance of seven aspects of retirement income policies to a representative sample of 1,066 New Zealanders, (ii) document the diversity of retirement income policy preferences in a systematically quantifiable manner, and (iii) rank three different retirement income policies from an individual well-being perspective. The ranking of policies can be interpreted within a well-being framework as the technique estimates how each person ranks each policy option in terms of their own preferences.

To estimate respondents' preferences we implemented an online survey using the software package 1000Minds. Rather than have people evaluate complex policy packages that affect multiple criteria simultaneously, this software more accurately estimates people's preferences by getting them to make a sequence of comparisons that only include two policy criteria at a time. Each comparison requires the respondent to reveal their willingness to trade-off an improvement in one criterion for a worsening of the other, and the software uses these responses to estimate each respondent's complete relative preference ranking over the seven criteria. The preference rankings over the separate criteria provide a way of estimating how people value complex policies, for example, whether their self-assessed well-being would be improved by a policy that increases income tax by 2 percentage points to fund an increase in the pension by $30 per week.

The seven criteria in the survey include the amount and the age of eligibility of New Zealand Superannuation (New Zealand's government retirement income scheme), the size of current and future taxes needed to pay for the scheme, whether or not the scheme should be universal or means-tested, and whether a compulsory saving scheme should be introduced instead of allowing people to save when and how they like. Each criterion has two categories that differ by amounts that, where appropriate, are broadly comparable in dollar terms. The survey was distributed to a representative sample by an independent sampling company, Colmar Brunton, in April 2014.

The distribution of responses enables us to explore whether there are some features of policies that most people think are relatively important, or relatively unimportant, or whether there are other features that are contentious. The retirement income policy criterion that is most important overall is universality – or more precisely, the absence of means-testing, which in this survey is an option to modestly reduce the weekly retirement income payment of people who have more than $200,000 in financial assets. The universality criterion is the most important criterion to 43% of the respondents, and has a mean rank of 3.15, on a scale from 1 (most important) to 7 (least important). The second

most important criterion concerns future tax rates. Most respondents think it is important that future generations avoid large tax increases and 65% would be willing to increase current tax rates by 2 percentage points if it meant tax rates on future generations would increase by 3 rather than 5 percentage points. The least important criterion, with a mean rank of 5.02, concerns saving flexibility. More than 50% of respondents consider the disadvantages of a 5% compulsory saving scheme to be sufficiently small relative to being able to save exactly when and how they like that they rank the 'saving flexibility' criterion as the one of the two least important. (This result may reflect that many people already save this amount.) The 'age of eligibility' criterion was the only criterion with a bimodal response; 35% of respondents indicate it is very important to keep the age of eligibility at 65 rather than increase it to 67, but a similar proportion indicate it is not important and would be willing to raise the age to achieve other objectives.

The difference between the mean ranks of the most and least important criteria is small because the public has diverse preferences over the relative importance of the seven criteria. The diversity of preferences can be measured by calculating the mean rank correlation of the preference rankings of all of the 1,066 respondents with each other. The mean rank correlation is only 0.08, not very different from the correlation coefficient of a sample whose preferences are randomly and uniformly distributed. The diversity of preferences about the relative importance of different retirement income policy features may help to understand why retirement income policy has been debated in New Zealand for nearly four decades.

The preference rankings vary with observable socio-demographic characteristics such as age, gender, and household income, but while the differences are statistically significant, they are not particularly large. Two socio-demographic factors stand out. Firstly, the preferences of people aged 65 and over differ from those aged less than 65 over five criteria: they think it is more important to have a higher pension and more wealth in retirement, they are more opposed to means-testing and less opposed to increases in current taxes, and they are less concerned to keep the age of eligibility at 65. Secondly, people living in low-income households, and people who are not confident they will be comfortable in retirement, have a stronger preference for keeping the age of eligibility at 65 than other groups; they are also more opposed to compulsion but less concerned about means-testing or future tax rates. New Zealanders with European ethnicity tend to be older and wealthier than New Zealanders with non-European ancestry, and are more in favour of raising the age of eligibility, more in favour of compulsion, and more opposed to means-testing than non-European New Zealanders. These results suggest that there is an element of self-interest in what people revealed in this survey, although in line with international evidence the effect of self-interest in small.

While preference differences based on age, income, gender and ethnicity are not particularly large, people can be sorted into five quite distinct preference groups or clusters reflecting five different average preference orderings. These clusters essentially reflect different attitudes, and can be labelled accordingly: there is a group that favours as little government intervention in retirement income policy as possible, for example, and another group that favours interventions that promote income redistribution. These clusters are primarily distinguished by whether their members give very high or very low ranks to three criteria – whether or not New Zealand Superannuation should be subject to means-testing, whether or not the age of eligibility should be increased from 65 to 67, and whether or not a compulsory saving scheme should be introduced. (While the size of future taxes is the second most important criterion on average, it is not a criterion that distinguishes people because most people are opposed to steep increases in taxes on

the next generation.) The members of the preference clusters are not strongly associated with particular socio-demographic characteristics. Consequently, the differences in the relative importance of different retirement income criteria primarily reflect differences in preferences, not differences in more observable characteristics such as age or income.

Given the wide diversity of preferences over retirement income policy criteria, can we identify or develop particular policies that would enhance or reduce the well-being of a large number of New Zealanders? To partially answer this question, we estimate how each of the 1,066 survey respondents would rank three different variants of New Zealand Superannuation: the current form; a variant in which the age of eligibility is increased by two years and taxes on future generations are reduced; and a variant in which the age of eligibility is unchanged but current taxes are increased by 2% of taxable income to reduce the tax increases facing future generations. Since the policies are ranked for each respondent using estimates of their individual preferences, the results have a natural interpretation as the way each of the retirement income policies affects the respondent's self-assessed well-being. Note that the respondents were not asked to rank these policies directly, but we infer their relative ranking from the way they answered the survey questions about the relative importance of different retirement income criteria.

The ranking exercise shows that, despite considerable diversity in preferences, a policy that increases current taxes to prevent larger tax increases on future generations is the most preferred policy for more than half of the population, and the least preferred policy for only a sixth. In contrast, a policy that raises the age of eligibility is the most preferred policy for only a sixth of the population and the least preferred policy for more than half. The current form of New Zealand Superannuation is the most preferred policy for a quarter of the population, and the least preferred policy of a similar fraction. These results suggest that a policy to more aggressively prefund New Zealand Superannuation would be viewed by a majority of New Zealanders as welfare enhancing, and by relatively few as welfare reducing. In contrast, a policy to increase the age of eligibility would improve welfare the most for only a relatively small number. As only three retirement income policies were ranked, not a complete set, we cannot conclude that a policy to more aggressively prefund New Zealand Superannuation is the best policy for New Zealand to adopt. Nevertheless, it was the highest ranked policy for all population subgroups based on income, age, ethnicity, education, and gender.

While the paper has specific findings that are relevant for retirement income policy, the bigger question concerns the potential usefulness of this approach to inform policy. Any survey has limitations that encompass factors such as the way questions are framed, the potential for ambiguity, omitted survey topics, and, with 1000Minds, the size and nature of the trade-offs respondents are asked to make. Whether these limitations outweigh the information obtained is open to debate on a case by case basis. Nonetheless, the results in this paper show considerable promise. It has been possible to demonstrate in a systematic and statistically rigorous manner that New Zealanders have considerable diversity in the relative importance they regard several aspects of retirement income policy. It has been possible to characterise how these preferences differ across different population subgroups, and to show that differences in attitudes tend to be considerably larger than differences across observable characteristics. And it has been possible to show that, notwithstanding this diversity, it is possible to find a policy that will enhance the well-being of a large number of people, as well as a policy that is disliked by large numbers, when evaluated in terms of people's own preferences. In short, it appears that this approach is effective in informing policy advice that is conditioned by a better understanding of society's preferences.

Table of Contents

List of Tables

List of Figures

A practical approach to well-being based policy development: what do New Zealanders want from their retirement income policies?

1   Introduction

The New Zealand Treasury's Living Standards Framework (LSF) specifies that the primary purpose of public policy is to enhance the capabilities and opportunities of individuals to pursue the lives they have reason to value. Although we do not know how individuals want to live, nor do we wish to pass judgement on how they should be living, numerous studies (such as those conducted by the OECD, the New Zealand Ministry of Social Development, and Statistics New Zealand) have identified a broadly consistent set of "domains of well-being", centred around economic, health, educational, safety, social and environmental considerations. The breadth of these well-being domains clearly suggests that the sources of human well-being are multi-dimensional and complementary in nature. Consequently, the LSF has deliberately adopted a broader, multi-dimensional, and integrated approach to economic, environmental and social policy advice that promotes wider well-being on a sustainable basis (Girol Karacaoglu 2015).

Public policy choices – such as the type of retirement income policy a country might adopt, or the rules it adopts to manage the environment – depend largely on two main factors: the various outcomes that different policy options deliver; and the preferences that people have over these outcomes. Since people in a society have diverse preferences, public policy advice should often be presented as a set of conditional recommendations: that policy A is better if people have one type of preference, for example, whereas policy B is better if they have other preferences. Conditional policy advice can be straightforward if the outcomes associated with different policies are easily identified and preferences over different outcomes are clearly understood. In practice, however, conditional policy advice is difficult to provide because policy interventions affect many different outcomes, and preferences over these outcomes are difficult to characterise and measure. For example, how can decision-makers know how many people favour a policy that increases the amount of the pension, if it simultaneously requires an increase in tax rates? And can they be sure that most of these people would not prefer a policy that raises the age of eligibility instead, because it reduces future tax rates?

An increasingly popular method of policy analysis that incorporates estimates of the public's preferences over different policy options is multi-criteria analysis (Belton and Stewart 2002). In this paper we apply a form of multi-criteria analysis that uses a sophisticated surveying methodology. This survey, which was developed by researchers from the New Zealand Treasury and the University of Otago in conjunction with the Commission for Financial Capability, is used to investigate the relative importance of several aspects of retirement income policies. The survey was conducted on a representative sample of 1,066 New Zealanders by an independent surveying firm, Colmar Brunton. Its primary purposes are to document the diversity of retirement income policy preferences in the population in a systematically quantifiable manner, and to ascertain if a measure of well-being based on these preferences can be used to help evaluate policy options.

The key feature of the survey is its use of the multi-criteria decision making software package, 1000Minds (Ombler and Hansen 2012). The software is designed to help people make complicated choices by making them compare specific features or criteria of the choice options two at a time. Using an online survey, respondents indicate their preferences over a dozen or so 'simple' alternatives each comparing two criteria. The software uses these responses to estimate each respondent's complete relative preference ranking over the various criteria. The methodology is related to the recent literature that develops well-being measures from estimates of the relative value that people place on different factors that make up well-being (Benjamin et al 2014).

The survey investigates the relative importance of seven different retirement income criteria such as the amount and the age of eligibility of New Zealand Superannuation (New Zealand's government retirement income scheme), the size of current and future taxes needed to pay for the scheme, and whether or not the scheme should be universal or means-tested. Each criterion has two categories that differ by amounts that, where appropriate, are broadly comparable in dollar terms. The survey was refined and pretested by Colmar Brunton before they conducted it in April 2014.

The estimated preference rankings provide a way of estimating the effect of various policies on an individual's self-assessed well-being: for example, whether they consider increasing income tax by 2% to fund an increase in the pension by $30 per week would improve their welfare. The distribution of responses enables policymakers to discover whether there are some features of policies that most people deem relatively important, or relatively unimportant, and whether there are other features that are contentious. In turn, this information can be used to investigate whether particular policies are likely to be widely welfare-enhancing or widely welfare-reducing, notwithstanding the diversity of individual preferences. We investigate the relative ranking of three variants of New Zealand Superannuation as a first-pass demonstration of our approach.

The paper is structured as follows. Section 2 summarises the background literature and describes the survey. Section 3 explains the methodological approach used to analyse the survey data. The survey results are presented in section 4, and the estimates are used to evaluate policy options in section 5.

2   The retirement income survey

2.1   Using public input in the policy making process

In the last two decades, the governments of most OECD countries have significantly changed the ways they incorporate public input into their decision-making and governing processes. This effort has taken many forms including formal summits, citizen juries, and 'Big Society' meetings in the United Kingdom, economic summits and consultative working groups in New Zealand, consultative task forces in Australia, and the creation of the Office of Public Engagement in the U.S.A (Lees-Marshment 2015). In addition, most governments now use focus groups and public opinion polling to better understand the issues facing their constituents and to develop and test their policy ideas (OECD 1998). The trend towards greater public input is sufficiently well established that the focus is now on the ways that it can be done most efficiently rather than whether it should be done at all (OECD 2001). Indeed, Lees-Marshment (2015) observes that many of the ways public input is incorporated into the policy process are ineffective because the information is not captured systematically and provided in a way that can inform decisions by Ministers.

One approach that uses public input to inform policy decisions in a more systematic way is multi-criteria analysis (Arrow and Raynaud 1986; Belton and Stewart, 2002). Originating in the operations research literature, this approach encompasses a range of techniques that are used to systematically analyse the relative importance of different aspects or criteria of complex problems (Renn et al 1993). The techniques range from the relatively informal to those based on complex software algorithms designed to identify the relative importance of different criteria to large numbers of different people (Devlin and Sussex 2011). These techniques have frequently been used by governments to improve the allocation of health expenditure, and to address environment, energy, and natural resource planning problems (Mendoza and Martins 2006; Gamper and Turcanu 2007; Devlin and Sussex 2011).

This paper uses a particular algorithm, PAPRIKA (Hansen and Ombler 2008), and a particular software package, 1000Minds (Ombler and Hansen 2012), to implement a multi-criteria analysis of retirement income policy. This software has been used by many Government agencies to help find solutions to complex 'micro-level' problems: for example, New Zealand's Ministry of Health uses it to help allocate elective surgery procedures based on an expert assessment of the extent that an intervention will improve different aspects of a patient's health, and the extent that the patient's well-being will be enhanced by these improvements (Devlin and Sussex 2011). In contrast, this paper uses the software package to systematically estimate the relative importance of different aspects of retirement income policy to a large representative sample of the general public. Obviously, the general public are frequently polled about their attitudes to policies and to government expenditure patterns in reasonably sophisticated ways. (See, for instance, OECD (1998) or the analysis and reviews of Wezlien (1995) and Soroka and Wezlien (2005)). Nonetheless, we believe this is the first large scale attempt by a government department to use decision-marking software to estimate the public's preferences over different features of policies in a manner that can be directly incorporated into the policy making process.

The approach we use is closely related to the bourgeoning literature that uses web-based multi-criteria surveys to establish the relative importance of the factors that improve well-being (for example Benjamin et al 2012, Benjamin et al 2014; OECD 2014). It differs from this literature, however, for two reasons. First, its focus is the way that several different aspects (or criteria) of a set of different policy options affect well-being. This is in stark contrast to the rest of the literature that examines how a range of different factors improves general well-being. Secondly, as we explain below, the focus is on the diversity of individual preferences and the ways individuals evaluate policies, rather than the average preferences of the whole population.

Globally, of course, there have been large numbers of surveys about attitudes towards retirement income policy. Many of these have asked people about the appropriate role of government in the provision of retirement income, about the appropriate amount of government retirement income, about the size of taxes, and about the extent that retirement incomes should be provided universally or on a means-tested basis. Some of these studies have asked questions that explicitly require respondents to make tradeoffs between one aspect of a policy and another, for example whether a respondent would be prepared to raise taxes to provide larger retirement incomes (e.g. Boeri et al (2002) for Italy and Germany; Van Els (2003) for The Netherlands; Evans and Kelley (2004) for Australia; or Fourati and O'Donoghue (2009) for Ireland). This study is clearly in this tradition, but uses a different survey technology, one we believe to have several advantages. First, the survey respondents are not asked for their attitudes about complete complex policy packages that affect multiple criteria simultaneously. Rather, they are asked about their preferences towards particular features of these policy packages, two criteria at a time. These comparisons should be much easier for respondents to make. Secondly, the technology enables respondents to compare a large set of options, not just a few select pairings. For example, we obtain information on the relative importance of the pension amount and the age of eligibility, the pension amount and future tax rates, and the pension amount and the amount of means-testing, not just one of these combinations. This enables us to estimate a full ranking of the relative importance of each criteria for each person, which previous papers have not been able to do. Thirdly, the technique allows an indirect estimate of each respondent's preferences over a large number of policy packages that is based on the value each respondent places on each of the features of packages, rather than just providing a ranking of the small number of complex options actually included in the survey. Thus the technique can be used to develop new policies, as well as evaluate those included in the survey.

2.2   The multi-criteria decision making approach

Multi-criteria decision analysis has been developed to assist individuals or groups to make complex choices over outcomes that involve multiple criteria or dimensions in an explicit, consistent and transparent way (Belton & Stewart 2002). There are several approaches, all of which identify a set of criteria that are used to evaluate an outcome, and then estimate the relative importance of each of these criteria. A typical multi-criteria decision making analysis has the following elements (Fülöp 2005).

1. An identification of the broad survey context – in the present context, a set of retirement income policies that generate different outcomes.

2. A set of relevant criteria by which the different policies will be ranked.

3. A process to estimate the relative importance of the criteria for members of the target population.

4. An evaluation of the different policies based on (i) an estimate of the effects of the different policies on each of the relevant criteria and (ii) the estimates of the importance of the criteria to the target population.

In the survey, each criterion is represented by a discrete list of possible outcomes or categories. The categories within each criterion are ranked from lowest to highest according to the benefits they provide a person. For example, if the categories for the 'age of eligibility' criterion were '65 years', '67 years', and '70 years', the category '67 years' would be ranked lower than '65 years' as a person receives a pension for fewer years.

This paper uses the PAPRIKA method (Potentially All Pairwise RanKings of all possible Alternatives) implemented through the 1000Minds software (Ombler & Hansen 2012) to estimate a respondent's preference ranking.[1] Respondents are presented with a series of hypothetical choices in an online survey, each of which involves scenarios that combine two criteria. In each case one of the combinations has a highly ranked category from one criterion and a lowly ranked category from the other, so that each selection indicates the relative importance of the categories to the respondent. Figure 1 is an example of a trade-off question from the survey. Respondents choose the combination of criteria they prefer from the two alternative scenarios: the one on the left retains the age of eligibility at 65 but requires current taxes to increase by 2%; the one on the right keeps current taxes the same, but raises the age of eligibility to 67. A respondent chooses his or her preferred combination, or indicates that they are indifferent between the two scenarios. Once the selection is made, the respondent is presented with another hypothetical scenario using categories from two randomly selected criteria.[2]

Figure 1 – Example of a trade-off question using the PAPRIKA scoring method

---

[1] The methodology is discussed by Hansen and Ombler (2008). While the PAPRIKA method and the 1000Minds software have been used by researchers in many countries and fields, this is the first time it has been implemented in New Zealand by a major surveying company.

[2] As the software randomly chooses the trade-off questions for each respondent, the first question seen by one decision-maker is unlikely to be the same as the first question seen by another. Changing the order of questions reduces or eliminates potential 'order biases' (Landon 1971, Perreault 1975, Dillman 2007).

Figure 2 is another example of a trade-off question from the survey. Respondents are now asked to choose which of the following two scenarios they prefer: the one on the left raises current taxes by 2% and taxes on the next generation (i.e. not you) by 3%; the one on the right has no change to current taxes and raises taxes on the next generation by 5%. The process is repeated until the algorithm has enough information to estimate a complete preference ranking over the criteria. At the end of the survey, respondents also provide some basic demographic and economic data to help analyse the survey results.

Figure 2 – Another example of a trade-off question

Although any number of criteria and/or categories can be included in the survey, the number of possible questions increases exponentially in the number of criteria and categories. The PAPRIKA method drastically reduces the number of choices that respondents have to make by automatically excluding 'dominant' pairwise comparisons and by using the property of transitivity to implicitly answer other questions. Nonetheless, to avoid overly long surveys, both the number of criteria and the number of categories needs to be selected sparingly.

A comment on the transitivity assumption is warranted. We can compare the responses of those who explicitly answered a particular scenario pair with those who were not faced with the same scenario pair owing to the transitivity assumption. If the fraction of people ranking one criterion higher than the other is similar for these groups, inference based on the transitivity principle would appear to be appropriate. We find little difference in the relative rankings of the two groups of respondents. Consider, for example, the relative importance of the 'current taxes' and the 'future taxes' criteria. 63% of the respondents were asked this question directly, of whom 66% ranked 'future taxes' to be more important than current taxes. The relative ranking imputed for the remaining 37% of respondents was 63%. The similarity of these two numbers in this example, and in other examples that we examined, suggests the transitivity assumption holds adequately in this survey.[3]

---

[3] The closeness of the relative ranking of a criteria pair for those who answered the questions directly and those for whom a ranking was imputed indirectly may be an increasing function of the difference in the average ranking of the two criteria. As the 'future taxes'
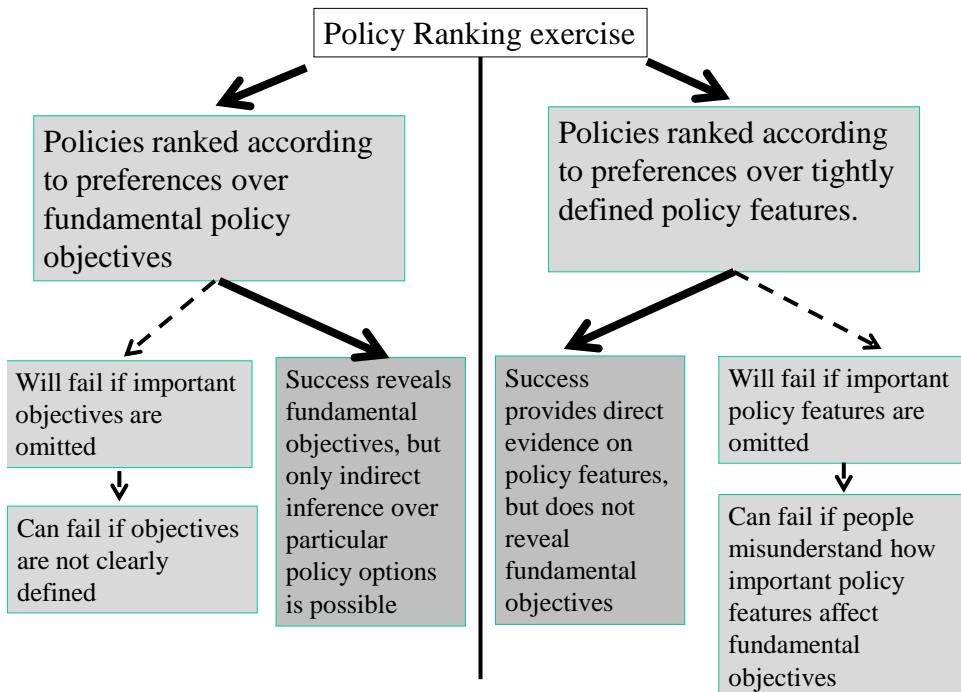
2.3   The survey strategy

The survey criteria were chosen after a lengthy process that involved an extensive review of the retirement income policy literature, the results of a previous retirement income survey conducted on a non-representative trial group of public servants, and discussions with several focus groups. The number of criteria and categories was kept to a minimum to reduce the size of the survey. The actual questions were extensively debated and trialled so that they were clear and concise.

There are two broad strategies for choosing criteria. The first is to find out respondents' preferences over the fundamental aspects of well-being as applicable to the objectives of retirement income policy. These objectives might include such aims as the minimisation of elderly poverty, or the provision of sufficient income to enable individuals to maintain their standard of living in retirement. This approach is adopted by several recent papers that attempt to measure the fundamental determinants of life satisfaction, such as Benjamin et al (2012, 2014). The second approach is to ascertain preferences over specific policy features that might help people achieve these fundamental objectives. These policy features could include the age of entitlement or the actual amount paid per week.

Figure 3 outlines some of the advantages and disadvantages of each type of strategy. The advantage of estimating preferences over fundamental policy objectives is that, if successful, the survey finds out information about the relative importance of the fundamental policy objectives. As well as allowing particular policy options to be ranked, this information can be used to create policy options that are aligned with the way people wish to live. In addition, the process of ranking different policies depends on experts' assessments about the ways particular policies affect the fundamental policy objectives, not the respondents' own assessments, which may be inaccurate. There are several disadvantages, however. If some of the fundamental objectives are excluded from the survey, the ranking of different policy options is likely to be wrong, and inference about specific policy features may be invalid. Even if they are included, it may be difficult to find criteria that are sufficiently clearly defined that reliable inference is possible, as people may interpret the criteria differently. In addition, even if the survey is well constructed, inference about the value of particular policies is indirect, and depends on the accuracy of the experts' assessment of the way policies affect the fundamental policy objectives.

---

criteria had an average ranking of 2, and the 'current taxes' criterion had an average ranking of 5, we might expect the algorithm to work well in these circumstances.

Figure 3 – Survey strategies

Policy Ranking exercise

Policies ranked according to preferences over fundamental policy objectives

Policies ranked according to preferences over tightly defined policy features.

Will fail if important objectives are omitted

Can fail if objectives are not clearly defined

Success reveals fundamental objectives, but only indirect inference over particular policy options is possible

Success provides direct evidence on policy features, but does not reveal fundamental objectives

Will fail if important policy features are omitted

Can fail if people misunderstand how important policy features affect fundamental objectives

When the survey criteria are specific policy features, some of these problems are overcome, although different problems can arise. The advantage of directly surveying respondents about specific policy features is that respondents will make their choices taking into account their fundamental policy objectives, so the problem of omitting some policy objectives is side-stepped. Moreover, the survey will provide direct evidence about the relative merits of different policy features. However, criteria representing specific policy features do not provide direct insight about respondents' preferences over fundamental policy objectives, making it difficult to devise new policy options. Moreover, responses combine preferences over fundamental objectives and participants' understanding of how different policy features affect these objectives. If their understanding is incorrect their reported preferences will reflect a misunderstanding of the way the world works. Finally, because not all policy features can be included in a survey, the ranking of complex policies may be unreliable because of the omission of some features.

We decided to survey people about tightly-defined policy features rather than the fundamental objectives of retirement policy. There were three reasons for this choice. First, our previous trial survey of public servants explored people's fundamental objectives. The length and complexity of this survey convinced us that it would be difficult to construct a survey about the fundamental objectives of retirement income policy that would be sufficiently short and clear for it to be answered by a wide cross-section of people. Secondly, one of the lessons from the wider literature analysing retirement income policy is that the ways many people view the objectives of retirement income policy depend on the mechanisms chosen to deliver the objectives; thus it can be difficult to establish preferences about objectives separately from preferences about policy options.[4] Thirdly, we wanted to be able to provide direct evidence on preferences about

---

[4] Bowles and Gintis (2000) and Fong (2001) show that many people have strong preferences over delivery mechanisms as well as outcomes: for instance a person may prefer a mandatory saving policy that delivers similar retirement incomes as a Government pension scheme because it requires people to save themselves rather than involves transfers to the "undeserving poor".

some policy reform options that were under discussion in New Zealand at the time of the survey. Although indirect evidence about specific policy choices such as the age of eligibility can be inferred from questions about fundamental objectives, these inferences will be invalid if some of the fundamental objectives that are important to people are omitted. Direct questions about aspects of policies do not suffer from this problem.

2.4   The survey criteria

The choice of criteria was motivated by the types of broad retirement policy options that were being discussed in New Zealand at the time of the survey. To understand this discussion, it is useful to note that government retirement policy options are typically classified three ways (see Figure 4). Tier 1 or universal schemes provide a retirement income funded from general taxation to eligible people irrespective of the amount they contribute during their working-age years. These incomes can be the same for all people (universal) or they can be means-tested. Tier 2 or contributory schemes provide a retirement income that increases with the amount people contribute during their working-age years. There are two basic types. In a tax-based contributory scheme, the government collects social security taxes when people are working and pays them a retirement income that depends on the amount they contribute over their lifetime. In a compulsory saving scheme, people are obligated to place a certain fraction of their labour income in a saving scheme, and the contributions, along with accumulated earnings, are available for their use upon retirement. Tier 3 schemes are voluntary, and typically encourage retirement saving by offering people subsidies or less punitive tax arrangements than ordinary saving. New Zealand has a universal tier 1 scheme, New Zealand Superannuation, that is largely funded on a pay-as-you-go basis from general taxation, and a subsidised tier 3 scheme, KiwiSaver, that was introduced in 2007.

Figure 4 – Types of retirement income policies
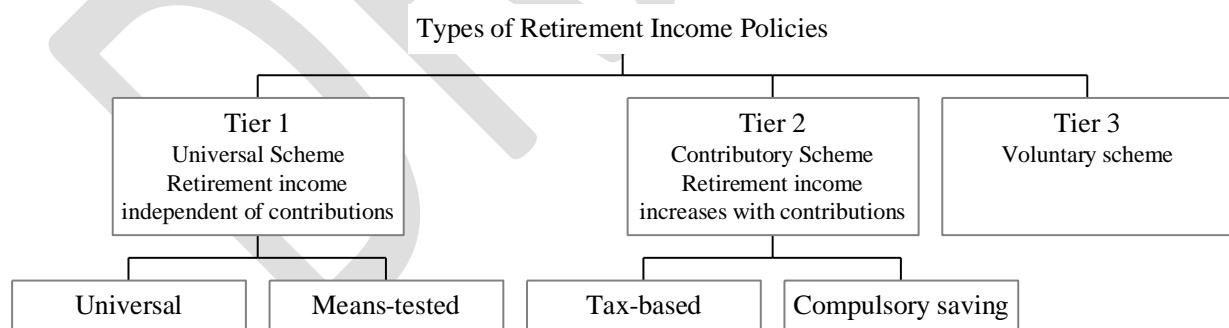
At the time of the survey, several reform options providing the background context for the choice of criteria we included were under public discussion. These reform options included:

1. maintaining the current form of New Zealand Superannuation, but raising the age of eligibility;

2. introducing a means-test for New Zealand Superannuation;

3. prefunding New Zealand Superannuation, by raising current taxes to prevent even larger tax increases in the future; and

4. introducing a compulsory saving scheme.

The criteria were chosen after a lengthy process involving consultation with members of the general public in several focus groups. The purpose of the focus groups was to identify all the relevant criteria for the survey. People from a wide range of backgrounds took part in the focus groups. These groups included students, retirees, women, Maori, Pacific Island people, disabled people, retirement policy experts, and representatives from Grey Power. The focus group meetings were held in Dunedin, Wellington and Auckland and consisted of attendees from around New Zealand. The meetings were structured to allow free-ranging discussions on retirement income. They began with a description of the retirement income policies of four unnamed countries, along with a summary of the amount of money that people in different circumstances could expect in each country. The policies included the universal scheme in New Zealand, the compulsory saving scheme in Australia, the taxed-based contributory scheme in the United States and a hypothetical retirement scheme. After the presentation, group members discussed why they preferred one policy over the others to help us uncover the relevant criteria to be included in our survey. It was made clear to group members that their individual and group preferences over the various retirement income policies would not be reported but would be used to help formulate the criteria in the survey.

The seven criteria we use concern the relative importance of several features of a universal government pension and a compulsory saving scheme (see Table 1). Five features of a universal pension scheme were included: the amount of the pension; the age of eligibility; whether or not it was means-tested; and the size and timing of taxes to pay for the pension. Two category levels were selected for each criterion. The category levels were selected so that the difference in the dollar value of the different categories for each criterion were broadly comparable so that people would be making choices over different design features worth similar amounts. For example, the two categories for the pension size differ by $30 per week, or by about $30,000 over a 20 year period. Similarly, the two categories for the age of eligibility differ by two years, or by about $30,000 of pension payments.

Table 1 – Important features of retirement schemes

Universal scheme	**Compulsory Saving Scheme**
The amount of the pension.	The desirability of accumulated savings.
The age of eligibility.	The importance of saving flexibility.
The desirability of means-testing.	
The willingness to increase current taxes to pay for the pension.	
The willingness to increase taxes on future generations to pay for the pension.	

The criteria concerning compulsory saving schemes were harder to formulate. In the end two were included. One criterion concerns whether people would find the lack of flexibility of a compulsory scheme inconvenient. The other concerns the value of having a larger quantity of assets available to save or spend upon retirement. A key feature of this

criterion was that the additional sum would be proportional to lifetime income, because if a compulsory saving scheme is adopted high income people will have more savings when they retire than low income people. One reason for choosing these criteria was to establish whether people preferred to have retirement schemes that provide identical pension payments to everyone or schemes that provided greater retirement incomes to those who saved more.

Table 2 lists the seven criteria. In addition to the different dimensions of retirement income policy that we wished to rank, we wanted (i) to avoid too much repetition; (ii) people to answer fewer than 15 questions; and (iii) people to be able to complete the survey in less than 10 minutes. The survey was pretested with some members of the focus groups, the staff of Colmar Brunton and the Commission for Financial Capability, and a sample of randomly selected New Zealanders provided by Colmar Brunton. Colmar Brunton also provided detailed feedback based on in-depth interviews with some participants. After this feedback, the questions were refined further.

Two categories were chosen for each criterion to reduce the number of questions each respondent would answer. The categories of the criteria were chosen with an eye to ensuring that the results of the survey could be used to evaluate different policies and that the differences between categories were broadly comparable in dollar terms. The three policies we chose to evaluate have been subject to extensive discussion in New Zealand and are: (i) maintaining New Zealand Superannuation in its contemporaneous form; (ii) raising the age of eligibility; and (iii) increasing taxes to partially prefund New Zealand Superannuation. For this reason, the baseline categories for the age of eligibility, the amount of the pension, and the means-testing regime correspond to the 2014 New Zealand Superannuation scheme. In turn, the baseline categories for the current and future tax criteria are the taxes that would be needed to fund New Zealand Superannuation now and in the future. The second category for the age of eligibility, 67 years, was chosen as it is an age that was often mentioned in contemporaneous public debate.[5] As two years of retirement income is approximately equivalent to $30 per week over the average length of time someone receives a pension, the second category of the pension amount was chosen to be $30 per week higher than the base category, so that the two criteria could be meaningfully compared.[6] Furthermore, the size of the tax increase necessary to support an increase in the size of the pension was calculated to be approximately 2% of personal income, and for symmetry we chose to vary current and future taxes by the same amount. Similar considerations made us choose the second category of the means-testing criterion so that the revenue raised by the means-test was approximately the same as the revenue saved by raising the age of eligibility.

The respondents typically answered twelve questions, and took five to ten minutes to answer the survey. To assuage concern that the respondents may not have understood the surveying technique, the software designers included a consistency test that required the respondents to repeat two of the comparison questions at the end of the survey. These comparison questions included one of the most preferred and one of the least preferred criteria, making it easy for someone doing the survey in good faith to answer, while discriminating against respondents who may not have understood the questions or

---

[5] In addition, the age of eligibility is schedule to increase to 67 in the United States and Australia, and to 68 in Great Britain.

[6] The second category had a $30 per week increase rather than decrease in the value of the pension partly because there was no contemporaneous public debate suggesting that the pension should be reduced. While we did not test this conjecture, we do not expect preferences over an increase and a decrease in the size of the pension to be symmetric. Rather we suspect people would be more likely to support an increase in taxes to prevent a decrease in the pension they would be to enable an increase in the pension.

who answered the questions in a random fashion. Eighty percent of respondents answered both repeated questions consistently, providing evidence that the survey procedure was well understood. Respondents who did not answer both additional questions consistently were excluded from the sample and additional people were surveyed. We also excluded people who answered the survey very quickly (in less than 10 seconds per question), as those answering very quickly were often inconsistent in their responses.

Table 2 – The survey criteria

	The criteria	Mean rank
1	**Amount of NZ Superannuation everyone receives** • $360 a week (current level) • increases by $30 a week to $390	4.09
2	**Age when NZ Superannuation starts** • 67 years (2 years later) • 65 years (current policy)	3.92
3	**Extra taxes to be paid now?** • everyone pays 2% more taxes (EXAMPLE: $20 more each week if earning $50,000) • no extra taxes	4.15
4	**Extra taxes the next generation (i.e. not you) has to pay** • 5% more taxes (EXAMPLE: $50 more each week if earning $50,000) • 3% more taxes (EXAMPLE: $30 more each week if earning $50,000)	3.41
5	**Will everyone receive the same amount of NZ Superannuation?** • No, people with retirement savings greater than $200,000 have their NZ Superannuation reduced by $60 per week • Yes, everyone gets the same NZ Superannuation	3.15
6	**The amount of your personal savings to spend or invest when you retire** • 2 years of your average annual income (don't worry how you get this amount) • 3 years of your average annual income (don't worry how you get this amount)	4.27
7	**Savings flexibility** • it is compulsory to save 5% of your income each week (EXAMPLE: $50 put aside each week if earning $50,000) • you can save when and how you like	5.02

There are dimensions of retirement income policies that were discussed in the focus groups but were not included in the survey. In order to minimise the length of the survey and make it as clear as possible (consistent with advice from Colmar Brunton), we chose not to include questions on eligibility criteria other than age, or on tax-funded contributory tier 2 schemes. An outside observer would probably be most surprised by the exclusion of the latter, since tax-funded tier 2 retirement income schemes are the most common schemes in OECD countries. Nonetheless, we chose not to ask questions about these schemes because few people in the focus groups could grasp these concepts and we

included questions about compulsory saving tier 2 schemes as an alternative. It was very clear that people in the focus groups had a much better understanding of compulsory saving schemes than contributory tax-based schemes. In addition, because people in the focus groups considered personal saving-based retirement income schemes to be very different than tax-based retirement income schemes, we thought it important to have a tier 2 scheme where income deductions were clearly identified as savings and not taxes.[7]

One issue that underlies the whole survey is framing. It is well known that the way questions are framed can have an enormous effect on survey responses. Some well known authors such as Bartels (2003) argue that framing effects may be sufficiently crucial to the design of a survey that they fundamentally undermine the use of all surveys as a source of useful input. We – and all other authors who conduct surveys - are not so extreme in our views. Nonetheless, it is possible that the relative ranking of the responses in part reflects the way the questions were framed, and that the answers might be different if they were framed differently. Unfortunately, we were unable to test this issue by running two differently framed versions of the survey, although this might be possible in future research. For this reason, all of the results of the survey should be subject to the generic survey warning that they might have been different if the questions were framed differently. This said, the survey questions were pre-tested and modified by the staff of Colmar Brunton to ensure that they were easily understood by respondents. This does not eliminate framing issues, but we hope it minimises them.

3   Measuring diversity: the methodological approach[8]

3.1   Measuring average preferences

Each respondent's survey response can be characterised as a vector listing the rank given to each of the seven criteria listed in Table 1 eg x_i = (2 1 6 4 3 5 7). Let X be a set describing the preferences of a subgroup of m respondents, X = {x_1, x_2, ….,x_m}. One measure of the average preferences of the subgroup is the mean preference vector of all members of the group:

$$\bar{x} = \frac{1}{m}\sum_{j=1}^{m} x_j \qquad (1)$$

This vector will not typically correspond to any individual's preference ranking. A second measure of average preferences, the distance minimising vector, is discussed below.

 An estimate of the difference in the preferences of two population subgroups, X_1 and X_2, is the difference in their mean preference vectors. We use the Wilcoxon-Mann-Whitney statistic to test whether the differences for individual criteria are statistically significant. This statistic tests the hypothesis that the distribution of the ranks given to a

---

[7] Many people expressed a view that if taxes were deducted from income to provide retirement incomes, then all recipients should receive the same retirement income, whereas if savings were deducted from incomes for retirement, people should have retirement assets in proportion to the amount they saved.

[8] For a comprehensive discussion of these metrics, see Mardin (1995).

particular criterion by each member of the group is the same for the two groups.[9] The hypothesis that the two groups have the same mean preference vector can also be tested for all criteria simultaneously using the Li and Schucany (1973) test, which is described below. The differences in the mean rank for each criterion are reported in Table 6 for various population subgroups.

3.2   Measures of diversity

The heterogeneity of a group of people can be measured by comparing the average distance between each member of the group with the average distance of a group of people in which all possible rank vectors are equally probable. (When there are 7 criteria the set of possible rank vectors, W, has 5,040 (=7 factorial) members, excluding ties.) Consider two people in the group X, x and y. The extent that they have similar responses can be measured by calculating the average 'distance' between the two vectors[10]:

$$d(x, y) = \sum_{i=1}^{7} (x_i - y_i)^2 \qquad (2)$$

The distance between two people with identical views is zero, whereas the distance between two people with diametrically opposed views is 112 when there are 7 preference criteria.[11]

Given the distance metric, diversity is measured in two ways. The first is the average distance between the members of the group and the distance minimising vector $\omega(X)$, which corresponds to the median preference vector of the group. The distance minimising vector $\omega(X)$ is the vector that has the minimum average distance to the m preference rank vectors in X:

$$\omega(X) = \arg\min_{\omega \in W} \left(\overline{d}(\omega, X) \right) = \arg\min_{\omega \in W} \left(\frac{1}{m} \sum_{j=1}^{m} d(\omega, x_j) \right) \qquad (3)$$

The average distance for a subgroup is compared to the average distance of the uniformly distributed preference group, that is, the group in which each of the 5,040 possible rank vectors is equally likely. (When there are 7 preference criteria, the average distance is $\delta_0 = 56$.) The normalised statistic τ compares the average distance between two groups to the average distance in a uniformly distributed sample:

$$\tau = 1 - \frac{\overline{d}(\omega(X), X)}{\delta_0} \qquad (4)$$

τ has a maximum value of 1 when the group is perfectly cohesive and is equal to zero when the group is diverse and preferences are uniformly distributed. The estimated standard deviation of τ is used to calculate confidence intervals for τ.

---

[9] A high absolute value of the test statistic indicates the distribution of the ranks given to a particular criterion is different for the members of the two groups, and thus that the difference in the means is statistically significant.

[10] For example if x = (2 1 6 4 3 5 7) and y = (6 2 1 4 5 3 7) the distance is $(2-6)^2 + (1-2)^2+(6-1)^2+(4-4)^2+(3-5)^2+(5-3)^2+(7-7)^2 = 50$.

[11] The maximum distance is for vectors (1 2 3 4 5 6 7) and (7 6 5 4 3 2 1) and permutations thereof.

The second measure of diversity is the mean Spearman rank correlation coefficient between all possible pairs in the group. The Spearman rank correlation between two preference vectors x and y, each with $n = 7$ criteria, is

$$\rho_{xy} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^{n} (x_i - y_i)^2 \qquad (5)$$

The mean Spearman rank correlation is calculated for all m(m+1)/2 possible pairs, and has a value between -1 and 1. A group that is uniformly distributed has a mean Spearman rank correlation equal to 0.

Comparing preferences between two groups

The mean distance and mean Spearman rank correlation measures can be used to test the hypothesis that the members of two different groups have the same distribution of preferences. For example, the Li-Frawley-Schucany tests use the diversity within each group as a basis to examine whether or not the difference in the mean rank vectors of the members of the two groups is zero (Li and Schucany (1973); Schucany and Frawley (1975)). These tests are related to the mean Spearman rank correlation between each member of the first group and each member of the second group. These tests are complementary to the Wilcoxon-Mann-Whitney statistical tests described above, which are used to ascertain if two different groups of people have the same distribution of preferences over a single criterion. If the Wilcoxon-Mann-Whitney test finds that the two groups have different preferences over at least one criterion, then the Li-Frawley-Schucany test should find that the groups differ in terms of their overall preferences.

3.3   Cluster Analysis[12]

A group can be partitioned into a set of subgroups or clusters whose members have similar preferences. Each person is allocated to the cluster that has the nearest mean preference vector; by construction, all members of the cluster will have greater affinity with each other's views than with members of other clusters. We find the partition that minimises the average distance between each person and his or her nearest cluster, and measure the cohesiveness of each cluster either as the mean Spearman rank correlation coefficient or as the mean distance between the cluster members and the cluster distance minimising vector. In this paper, we find that the general public can be grouped into five clusters and we calculate the fraction of the group in each cluster. A multinomial distribution test can be used to test the hypothesis that the allocation of members of two subgroups across the clusters is the same.

A version of the Lloyds algorithm is used to find the k distance minimising clusters (Lloyd 1982). This algorithm begins with k random vectors and creates initial estimates of the k clusters by allocating each group member to the nearest vector. The mean of each cluster is then calculated, and the group members are reallocated to their nearest cluster, with the process repeated until a partition is found in which each person is allocated to his or her closest cluster. While this algorithm finds a local distance-minimising partition, it is not guaranteed to find the global distance minimising partition. Consequently, the process is repeated using 500 different sets of initial random vectors to find a partition close to the global minimum distance partition. While the algorithm will find a global minimum partition

---

[12] See Jain and Dubes (1988: chapter 3) for a comprehensive discussion about the application of cluster analysis to survey data.

when there are large numbers in the group, if the preferences of group members are widely dispersed and if there are small numbers of people (where 'small' includes our sample of 1066 people), partitions corresponding to different initial conditions can have quite different numbers of members even though they have similar average distances. As such, there can be considerable uncertainty in the precise location of the clusters.[13] To avoid being misleadingly-precise, we report the mean and standard-deviation of the numbers of people in each cluster for the third of the partitions with the smallest average distances.

4   Results

4.1   The sample

Table 3 provides information about selected economic and demographic characteristics of the survey respondents. The respondents are a representative cross-section of New Zealanders. The basic demographic characteristics are as follows: 46% of the respondents are male; 26% are aged less than 35, 59% are aged 35 – 64, and 15% are over 65; 24% are single; and 60% have children. In terms of ethnicity 76% are New Zealand European, 12% are Maori, 6% are Pacific Island people, and 12% are Chinese, Indian, or other Asian. Working patterns and household incomes vary across the sample: 53% of respondents work full-time, 17% worked part-time, and 15% are retired; 32% of respondents live in households with less than $50,000 income, 41% live in households with between $50,000 and $100,000 and 27% live in households with more than $100,000 income. In terms of qualifications 48% of the respondents have a degree, but 32% of the sample do not have post-secondary school qualifications. Respondents were also asked whether they were confident that they would be comfortable in retirement. Four options were offered: 'Not confident at all,' 'Not too confident,' 'Somewhat confident' and 'Very confident'. Nine percent of the sample said they were "Not confident at all," while 16% said they were "Very confident". 67% of respondents belong to KiwiSaver.

Table 3 – Selected economic and demographic characteristics of the survey respondents

Gender	Male 46%	Female 54%	
Age	< 35 years 26%	35 – 64 years 59%	65 years + 15%
Ethnicity	European 76%	Maori 12%	Other 12%
Children	Have children 60%	No children 40%	

---

[13] If all people can be tightly fitted into different clusters – that is, if people can be partitioned into groups comprising people with views that are very similar to each other, but quite different to other groups - the algorithm consistently finds the same clusters irrespective of the initial conditions and there is little uncertainty in the cluster estimates. If some people have widely dispersed views, their distance to the nearest and the second nearest cluster are similar, and so their membership of a particular cluster will depend on precisely who else belongs to that cluster, which determines the cluster sample mean. In these circumstances there is genuine uncertainty as to the exact centroids of the clusters, and therefore their precise membership.

Household Income	< $50,000	$50,000 – 100,000	$100,000+
	32%	41%	27%
Employment status	Full-time	Part-time	Retired
	53%	17%	15%
KiwiSaver membership	Yes	No	
	67%	32%	
Geographic Spread	Auckland	Other North Island	South Island
	33%	41%	26%

The results presented below have been reweighted to take account of the difference between the socio-demographic characteristics in the sample and the socio-demographic characteristics in the country, using weights provided by Colmar Brunton. The reweighting has little effect on the results as the sample is broadly representative.

4.2 The average level and dispersion of preference ranks: the full sample

Table 4 shows various measures of the average level and dispersion of the estimated preference ranks for the full sample. (Rank '1' means the criterion is important; rank '7' means it is unimportant.) There are three key results.

Table 4 – Average retirement income preferences in New Zealand

	Criterion 1 Pension Amount	Criterion 2 Age 65 / 67	Criterion 3 Current taxes	Criterion 4 Future taxes	Criterion 5 Means tests	Criterion 6 Wealth amount	Criterion 7 Flexible savings
	Whole sample n = 1,066						
Distance min vector	4	3	5	2	1	6	7
Mean rank	4.09	3.92	4.15	3.41	3.15	4.27	5.02
% rank 1	7.8%	21.2%	5.9%	12.4%	41.7%	5.4%	9.8%
% rank 7	10.7%	15.1%	5.3%	2.6%	14.3%	8.8%	38.1%
Coherence / Dispersion	Mean distance 40.4		τ 0.28	(sd) (0.012)	Mean Spearman Correlation 0.080		

The first result concerns the diversity of preferences. The mean distance between respondents' preference rank vectors is 40.4, and the mean Spearman rank correlation is 0.08. Although both measures are statistically different at the five percent significance level from the levels that would occur if people had uniformly distributed preferences, both measures indicate that New Zealanders have very diverse preferences about what they want from retirement income policies. This dispersion is the reason why the mean preference ranks for the different criteria have a narrow range, from a minimum of 3.15 to a maximum of 5.02.

The second result concerns the overall importance of the different criteria. The two highest ranked criteria concern (i) universality/mean-testing and (ii) future taxes rates. On average, respondents expressed a strong preference for universal rather than means-tested pensions, and are opposed to policies that result in steep increases in taxes on future generations. The lowest ranked criterion is the flexible saving/ compulsory saving criterion: few people thought there was much advantage from being able to save when and how they liked rather than being forced to join a compulsory saving scheme. In between, the other four criteria had mean ranks varying from 3.92 to 4.27. Three of these criteria - current tax levels, the amount of the pension, and the amount of wealth people have in retirement – were of moderate importance to most people. The fourth, the age of eligibility, had a bimodal distribution. The relative importance of the criteria is the same whether the distance minimising vector or the mean rank vector are the measures.

The third result concerns the distribution of rank preferences for each criterion. Table 4 also shows the fraction of the population who ranked each criterion either highest or lowest, and figures 4a–4f show the entire distributions of the responses.

Figure 5 shows that the 'universality/mean-testing' criterion is the most important criterion to the largest number of people. 42% of respondents rank universality rather than means-testing as the most important feature of retirement income policy, and an additional 20% rank it as the second or third most important. Only 23% of respondents ranked universality as 6th or 7th most important, implicitly indicating support for means-testing.

Figure 5 – Criterion that is important to most people 'Universality / no means-testing'

Percentage of respondents

No means-tesing / universal pension

Rank

Key: Rank 1 = highest rank / most important criterion
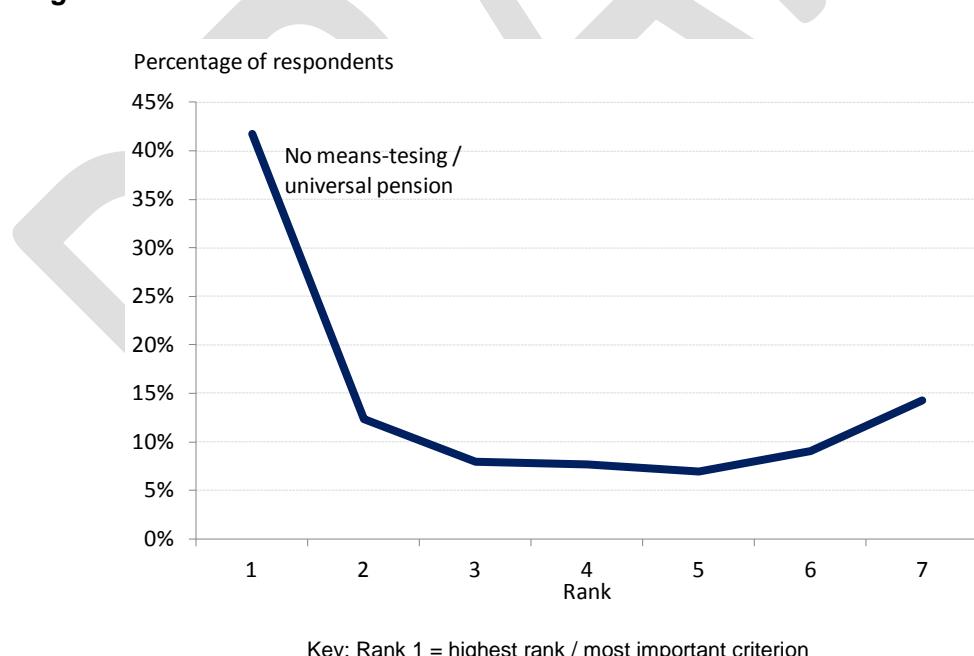
Figure 6 shows the distribution of preference ranks for the two tax criteria. Both are humped shaped: they are neither the most important criteria nor the least important criteria for most people, but are moderately important. The figure shows most people are more opposed to future tax increases than current tax increases: indeed, 65% of all respondents gave (low) 'future taxes' a higher weight than 'current taxes', while only 30% ranked them the other way around. (This split held for all population subgroups.) These responses strongly suggest that there is widespread opposition to the adoption of policies that impose high costs on future generations.[14]

Figure 6 – Criterion that is moderately important to most people 'Future taxes'

Key: Rank 1 = highest rank / most important criteria
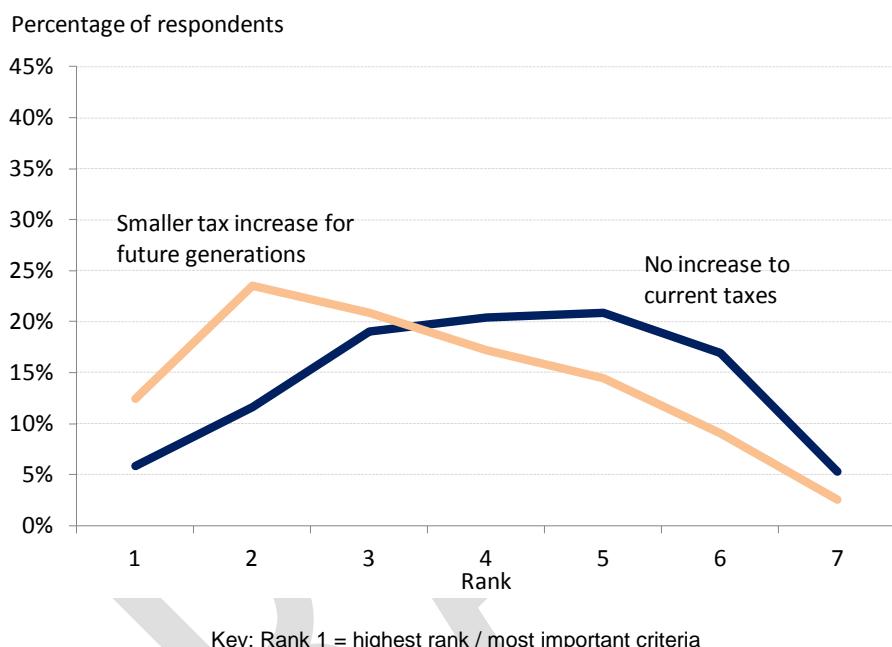
Figures 7 and 8 show the distribution of preference ranks for two other 'hump-shaped' criteria: the benefit of higher pensions, and the benefit of higher retirement savings. These are neither the most important nor the least important criteria to many people, but they are moderately important as can be expected: most people would like a higher pension or greater retirement savings. Overall respondents were nearly equally split as to whether they preferred a higher pension or higher retirement savings.

---

[14] Note that the survey criterion on future taxes explicitly refers to the effect on future generations, not themselves. If younger respondents answered the question thinking about their own future tax rates, for them the response is consistent with tax smoothing. This does not detract from our finding that people would be prepared to raise taxes immediately to reduce the rate of future tax increases; it merely changes the motivation for why they respond in this manner.

Figure 7 – Criterion that is moderately important to most people 'Size of pension'

Percentage of respondents

Key: Rank 1 = highest rank / most important criterion
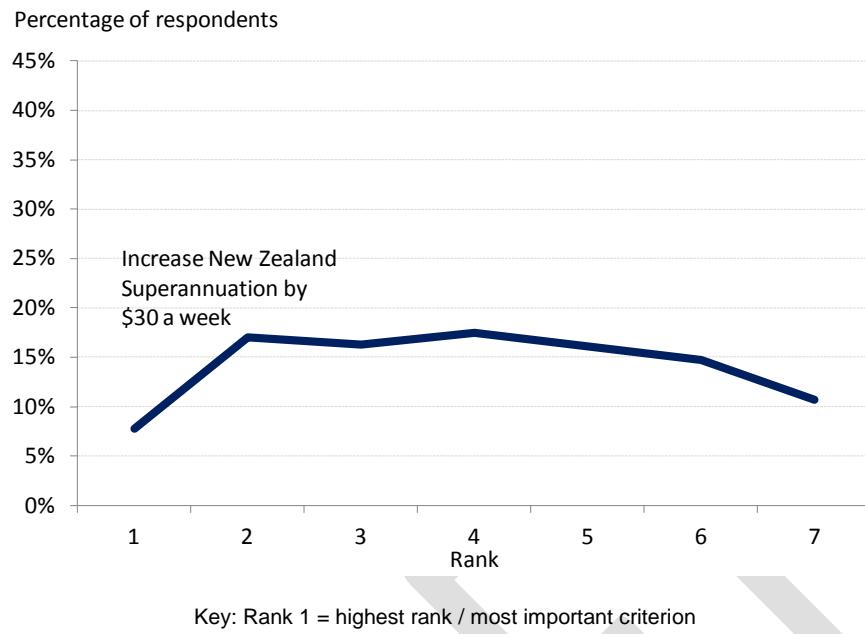
Figure 8 – Criterion that is moderately important to most people 'Size of retirement savings'

Percentage of respondents

Key: Rank 1 = highest rank / most important criterion
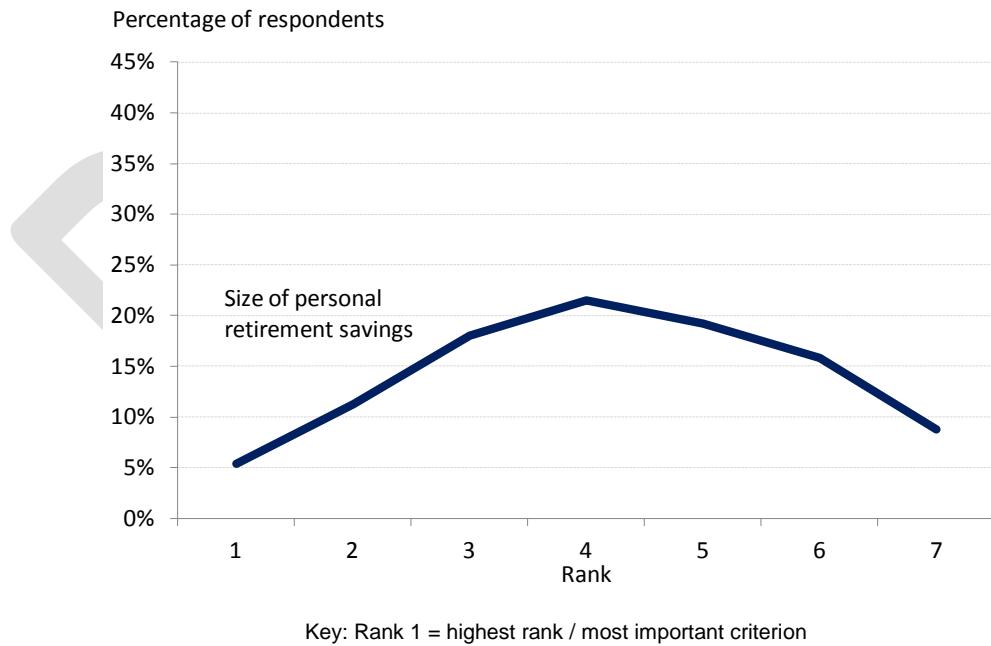
Figure 9 shows the distribution of preference ranks for the flexible saving / compulsory saving criterion. It has the opposite shape to 4: there are relatively few people who think saving flexibility provides large benefits, and many people who think it provides very few benefits. Overall, 38% of respondents indicated that saving flexibility was the least important of all seven criteria, and only 19% indicated it was one of the two most important criteria. This result suggests there would be little opposition to a compulsory saving scheme if it raised the amount of wealth available at retirement, possibly because many people already save this amount.

Figure 9 – Criterion that is unimportant to most people 'Saving flexibility / Compulsion'

Percentage of respondents

Key: Rank 1 = highest rank / most important criterion
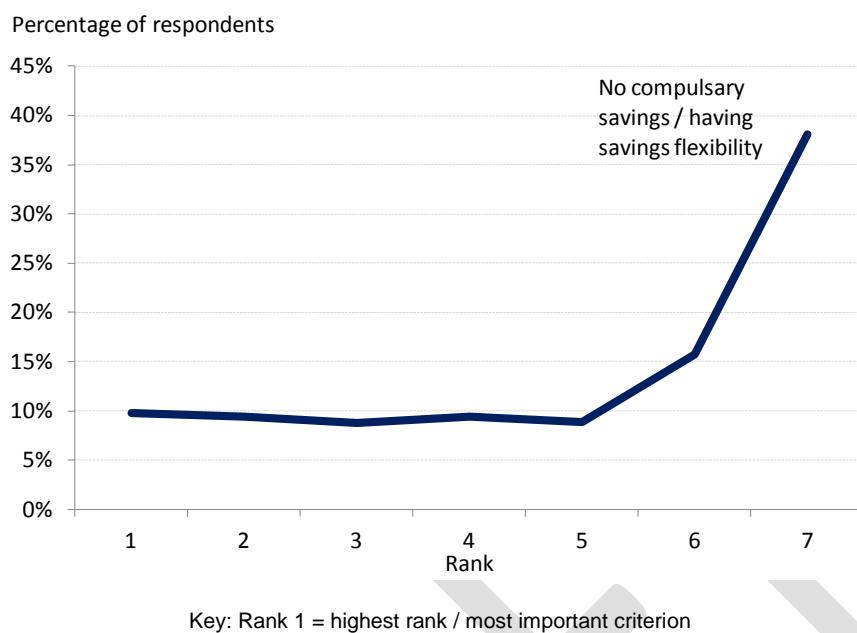
Lastly, Figure 10 shows the distribution of preference ranks for age of eligibility criterion. It is the only criterion with a bimodal response. 37% of the respondents indicate it is very important to keep the age of eligibility at 65 (1st or 2nd ranking), and 33% of respondents indicate it is unimportant (6th or 7th ranking). The criterion is important to people from low income households, to New Zealanders of Pacific Island ethnicity, and to those who are not confident about their retirement prospects, but unimportant to people over 65, to New Zealanders of European ethnicity, and to people who are confident about their retirement prospects. One rationale for this response that was frequently expressed in the focus groups was that low income people may disproportionately have manual jobs, and be less able to participate in the labour force after age 65 than people with less physically demanding jobs. A second rationale is that for some low-income people the pension is similar to, or more than what they currently receive from working.

Figure 10 – Criterion that is bimodal 'Age of eligibility'

Percentage of respondents

Key: Rank 1 = highest rank / most important criterion
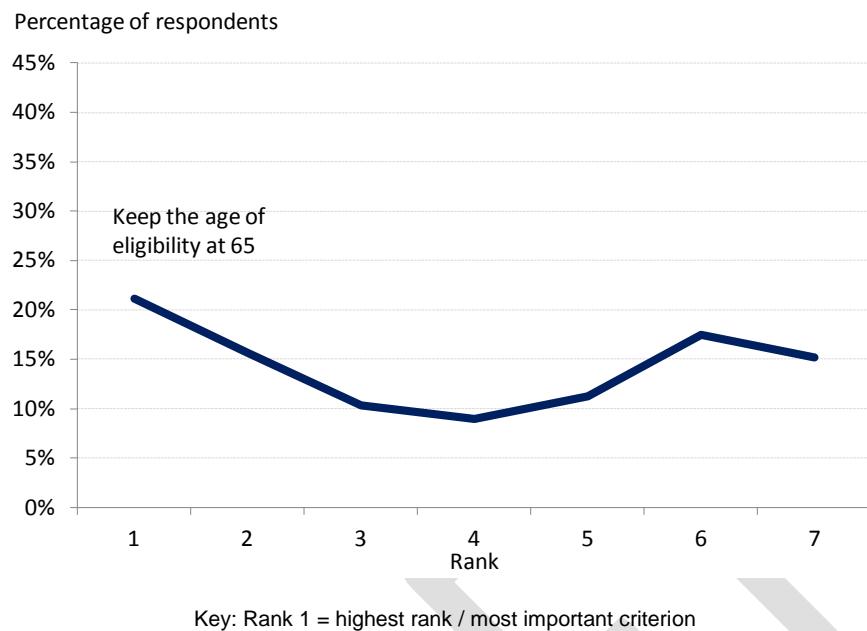
4.3   Results for population subgroups

It is natural to ask whether different population subgroups have different preferences. The short answer is that they do, but that these differences, while statistically significant, tend to be small. With one exception (the subgroup of Pacific Island people, discussed further below), there were few criteria where the mean preference ranks for population subgroups defined in terms of observable characteristics such as age, gender, household income, education or ethnicity differed by more than 0.5 ranks on a scale of 1 – 7. Indeed, the largest differences between subgroups occurred for groups that self-identified in terms of their expected comfort in retirement rather than for groups that could be identified in terms of measurable characteristics.

The analysis of population subgroups is conducted in two ways. First, we divided the population into subgroups and compared the mean preference ranks for a particular subgroup with all people not in that group: for example, people aged 65 or more versus people aged less than 65. In each case we calculated the mean rank of each criterion for the two subgroups, and used a Wilcoxon-Mann-Whitney test to test whether the distributions were the same.[15]   Table 5 provides the full comparison for people aged more than and less than 65 and Table 6 shows the mean differences for each criterion for various other subgroups. Obviously these subgroup comparisons do not condition on the other factors that may vary within each subgroup, and thus do not estimate the marginal effect of a socio-demographic factor on preference ranks. To do this we estimate a fractional multinomial logit model using the entire set of socio-demographic variables as independent variables. This model takes into account the loss of one degree of freedom that occurs when objects are ranked, or equivalently, that the sum of the relative ranks of seven criteria must equal 28. To estimate the model, the rank vectors are first converted

---

[15] We also calculated the Li-Schucany test statistic of the hypothesis that the two subgroups have the same mean vector of preferences across all criterion: this was rejected at the 1% level for all groups, and is not reported.

into a set of normalised weights that sum to one.[16] These weights are simultaneously regressed against dummy variables corresponding to each of the socio-demographic variable categories. The regression coefficients indicate how the weight of each criterion depends at the margin on each socio-demographic variable. For example, in Table 7 the estimated coefficient for people over 65 on the 'pension size' criterion is ─0.033, indicating that the mean preference rank for people over 65 is 0.9 (=0.033*28) lower (more important) than for people under 35.

Table 5 – Mean preferences of people aged more than and less than 65

	Criterion 1 Pension Amount	Criterion 2 Age 65 / 67	Criterion 3 Current taxes	Criterion 4 Future taxes	Criterion 5 Means tests	Criterion 6 Wealth amount	Criterion 7 Flexible savings
	People aged 65 or more n = 202						
Distance min vector	3	5	6	2	1	4	7
Mean rank	3.82	4.39	4.46	3.53	2.94	3.93	4.93
	Mean distance 38.7		τ 0.31	(sd) (0.028)	Mean Spearman Correlation 0.09		
	People aged 64 or less n = 864						
Distance min vector	5	3	4	2	1	6	7
Mean rank	4.16	3.81	4.08	3.38	3.20	4.35	5.04
	Mean distance 39.9		τ 0.29	(sd) (0.014)	Mean Spearman Correlation 0.08		
	Differences						
Mean difference	-0.34*	0.58**	0.38**	0.15	-0.25*	-0.42**	-0.11
WMW statistic	2.54	3.45	3.18	1.04	2.03	3.42	0.57
Significance	0.006	0.000	0.001	0.150	0.021	0.000	0.284
	Li-Schucany test that the groups have the same mean vector L* = 74.3 (0.000)						

---

[16] Each weight is equal to the relative rank divided by 28. The weights sum to 1, and each increase in rank is equal to a 0.036 change in the weight.

Table 6 – Mean difference by criteria for selected population subgroups

	N	Criterion 1 Pension Amount	Criterion 2 Age 65 /67	Criterion 3 Current taxes	Criterion 4 Future taxes	Criterion 5 Means tests	Criterion 6 Wealth amount	Criterion 7 Flexible savings
All People	1,066	4.09	3.92	4.15	3.41	3.15	4.23	5.02
Demographic characteristics								
Males	510	-0.23*	0.21	-0.07	0.14	-0.13	-0.19	0.26
Females	556	0.23*	-0.21	0.07	-0.14	0.13	0.19	-0.26
Single	253	-0.10	0.05	-0.10	-0.07	0.29	-0.14	0.07
Have children	624	0.13	-0.23	0.02	0.10	0.13	0.08	-0.23
Age <35	308	0.32**	-0.05	-0.43**	-0.20	0.24	0.19	-0.07
Age 35-64	556	-0.06	-0.32*	0.12	0.07	-0.04	0.10	0.13
Age 65+	202	-0.34*	0.58**	0.38**	0.15	-0.25*	-0.42**	-0.11
Region or ethnicity								
Auckland	353	-0.05	0.00	-0.18	0.01	-0.02	-0.04	0.27*
European	783	-0.16	0.39**	0.42**	-0.07	-0.42**	-0.28**	0.12
Maori	160	0.06	-0.20	0.07	-0.18	0.65**	0.05	-0.44*
Pacific Island	70	0.52**	-0.60**	-0.50**	-0.51**	0.80**	0.12	0.18
Asian	128	0.06	-0.29	-0.69**	0.36*	0.30	0.37**	-0.13
Highest education, and employment status								
Secondary school	340	-0.03	-0.13	0.19	0.19	-0.15	0.31**	-0.39*
Degree	466	0	0.30*	-0.22*	-0.16	0.11	-0.17	0.14
Full-time job	558	0.10	-0.02	-0.18	-0.04	-0.28	0.11	0.30*
Household income								
<$50,000	368	-0.21	-0.30*	0.06	0.27*	0.42**	0.08	-0.31**
$50-100,000	426	0.06	0.05	0.04	-0.25*	-0.02	0.07	0.04
>$100,000	272	0.17	0.29	-0.12	0.00	-0.48**	-0.18	0.32*
Confidence about adequacy of retirement income/ KiwiSaver membership								
Not confident	86	-0.58**	-0.67**	0.11	0.39*	1.14**	-0.15	-0.25
Very confident	183	0.43**	0.64**	-0.37**	-0.17	-0.43*	0.30*	-0.41
KiwiSaver	701	0.03	-0.22	-0.13	-0.22	-0.30	-0.02	0.85**
No KiwiSaver	365	-0.03	0.22	0.13	0.22	0.30	0.02	-0.85**

Each statistic is the difference in the mean preference rank for the identified subgroup with all people not in that subgroup. A negative number means the mean rank is lower (more important) for the subgroup. A * (**) indicates the hypothesis that the two groups have the same distribution of preferences can be rejected at the 5% (1%) significance level, using a Wilcoxon-Mann-Whitney non parametric test.

Table 7 – Fractional multinomial logit estimates of the effect of socio-demographic variables on the relative ranking of different policy criteria

Demographic Characteristics	Amount of NZS	Age of NZS	Current Taxes	Future Taxes	Universal / Means-Testing	Wealth Amount	Compulsary Savings
Upper North Island excluding Auckland	0.004	0.002	0.006	0.002	-0.008	0.004	-0.009
(ref group: Auckland)	(0.006)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)	(0.007)
Lower North Island	0.006	0.002	-0.001	-0.006	0.002	-0.003	0.000
(ref group: Auckland)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)	(0.005)	(0.006)
South Island	-0.006	-0.009	0.003	0.006	0.008	0.002	-0.003
(ref group: Auckland)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)	(0.005)	(0.006)
Male	-0.007	0.007	-0.003	0.004	0.000	-0.005	0.004
(ref group: Female)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)
35-64 years	-0.013	-0.005	0.012	0.007	0.005	-0.008	0.002
(ref group: <35 years)	(0.005)*	(0.006)	(0.005)*	(0.005)	(0.007)	(0.005)	(0.006)
65+ years	-0.033	0.030	0.028	0.003	-0.008	-0.036	0.016
(ref group: <35 years)	(0.009)*	(0.012)*	(0.009)*	(0.009)	(0.012)	(0.008)*	(0.011)
NZ European and Maori	0.010	0.009	-0.004	-0.013	0.019	-0.015	-0.007
(ref group: NZ European only)	(0.009)	(0.011)	(0.007)	(0.007)	(0.011)	(0.008)	(0.011)
Maori only	0.005	-0.011	0.003	0.002	0.002	0.012	-0.013
(ref group: NZ European only)	(0.011)	(0.012)	(0.008)	(0.008)	(0.012)	(0.008)	(0.012)
Chinese only	-0.002	-0.003	-0.020	0.006	0.003	0.011	0.005
(ref group: NZ European only)	(0.01)	(0.011)	(0.009)*	(0.009)	(0.011)	(0.009)	(0.01)
Pacific Island only	0.015	0.006	-0.025	-0.011	0.007	0.003	0.004
(ref group: NZ European only)	(0.011)	(0.013)	(0.01)*	(0.009)	(0.013)	(0.01)	(0.013)
Other	0.004	-0.027	-0.009	0.009	0.018	0.005	0.000
(ref group: NZ European only)	(0.006)	(0.007)*	(0.006)	(0.006)	(0.008)*	(0.006)	(0.007)
Other post secondary school qualification	-0.002	0.003	-0.001	-0.006	0.003	-0.013	0.015
(ref group: No to School Qualifications only)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)*	(0.006)*
University degree or equivalent	-0.001	0.013	-0.002	-0.008	0.010	-0.015	0.002
(ref group: No to School Qualifications only)	(0.005)	(0.006)*	(0.004)	(0.004)	(0.006)	(0.005)*	(0.006)
Not working	0.001	0.005	0.001	0.002	0.003	-0.011	-0.001
(ref group: Working)	(0.006)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)*	(0.007)
Retired	0.010	-0.006	-0.013	0.003	-0.010	0.006	0.010
(ref group: Working)	(0.01)	(0.01)	(0.008)	(0.008)	(0.011)	(0.008)	(0.011)
$30,001 - $50,000	0.001	0.009	-0.007	-0.002	0.000	-0.003	0.003
(ref group: $30,000 or less)	(0.008)	(0.01)	(0.006)	(0.006)	(0.008)	(0.007)	(0.009)
$50,001 - $70,000	0.000	0.020	0.000	-0.009	-0.010	-0.009	0.007
(ref group: $30,000 or less)	(0.008)	(0.01)*	(0.006)	(0.006)	(0.008)	(0.007)	(0.009)
$70,001 - $100,000	0.002	0.022	-0.003	-0.009	-0.016	-0.014	0.019
(ref group: $30,000 or less)	(0.008)	(0.01)*	(0.007)	(0.006)	(0.008)*	(0.006)*	(0.009)*
$100,001 or more	0.003	0.025	-0.004	-0.003	-0.023	-0.019	0.022
(ref group: $30,000 or less)	(0.008)	(0.011)*	(0.007)	(0.007)	(0.008)*	(0.007)*	(0.01)*
De facto relationship	-0.005	0.003	-0.002	0.004	0.001	-0.001	0.000
(ref group: Married or Civil Union)	(0.006)	(0.007)	(0.005)	(0.005)	(0.007)	(0.006)	(0.007)
Single	-0.004	0.011	-0.002	-0.002	0.002	-0.007	0.004
(ref group: Married or Civil Union)	(0.006)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)	(0.006)
Has children	0.010	-0.004	-0.002	-0.001	-0.001	0.005	-0.008
(ref group: No children)	(0.005)*	(0.006)	(0.004)	(0.004)	(0.006)	(0.004)	(0.005)
Somewhat confident	-0.008	-0.010	0.008	0.004	0.011	-0.013	0.008
(ref group: Very confident)	(0.005)	(0.007)	(0.006)	(0.005)	(0.008)	(0.005)*	(0.007)
Not too confident	-0.025	-0.008	0.019	0.003	0.015	-0.018	0.014
(ref group: Very confident)	(0.006)*	(0.007)	(0.006)*	(0.006)	(0.009)	(0.006)*	(0.008)
Not confident at all	-0.030	-0.030	0.013	0.015	0.036	-0.018	0.012
(ref group: Very confident)	(0.008)*	(0.01)*	(0.008)	(0.008)	(0.012)*	(0.008)*	(0.011)
In KiwiSaver	-0.007	-0.005	0.001	-0.004	-0.009	-0.008	0.031
(ref group: No KiwiSaver)	(0.005)	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)	(0.008)*
Has private scheme	-0.003	0.003	0.004	-0.002	-0.016	-0.002	0.014
(ref group: No private scheme)	(0.005)	(0.006)	(0.004)	(0.004)	(0.006)*	(0.004)	(0.006)*

The dependent variable is the criterion weight equal to the relative rank/28. A negative coefficient means that people with the particular socio-demographic variable consider the criterion to be important.

The results of the two analytical approaches are largely consistent. Many of the socio-demographic variables included in the fractional multinomial logit model are associated with statistically significant changes in the mean rank of one criterion or another, but the effects tend to be small, less than one rank. If a socio-demographic variable has a statistically significant effect in the fractional multinomial logit model, the associated population subgroup typically has a mean rank vector that differs from the rest of the population by a statistically significant amount, but again the differences are small, normally less than 0.5 ranks.[17]

The results can be analysed in terms of the particular criteria that are important to particular socio-demographic groups (i.e. an analysis of the rows of tables 6 and 7), or in terms of the socio-demographic characteristics that are associated with people who have strong views about each of the seven criteria (i.e. an analysis of the columns of tables 6 and 7). Considering the rows first, three distinctive features associated with different socio-demographic factors stand out. Firstly, the responses of people aged 65 and over differ from responses of those aged less than 65 across five criteria: it is more important to them to have a higher pension and more wealth in retirement, they are more opposed to means-testing and less opposed to increases in current taxes, and are less concerned to keep the age of eligibility at 65. As we discuss below, these results are consistent with people responding to the survey in a self-interested manner, although the size of the effects is small.[18]

Secondly, people living in low-income households have a stronger preference for keeping the age of eligibility at 65 than other groups; they are also more opposed to compulsion but less concerned about means-testing or future tax rates. The same preferences are shown by people who are not confident they will be comfortable in retirement, but they are more strongly held. (The latter group also expresses a much greater willingness to impose means-tests and would also like to see the size of the pension increased.) These responses are also consistent with self-interested behaviour.

Thirdly, the results in Table 6 suggest there are differences in the preferences of different ethnic groups. New Zealanders with non-European ethnicity tend to be more concerned to keep the age of eligibility at 65, are more opposed to increases in current taxes, and are more supportive of means-tested pensions than New Zealanders with European ethnicity. The differences are most marked for Pacific Island people. However, it appears that these differences mainly reflect the different age and income characteristics of non-European New Zealanders, as ethnicity is not an important factor in the fractional multinomial logit regressions. Once age and income are taken into account in the fractional multinomial regressions, the results in Table 6 that indicate Pacific Island people have relatively strong preferences for higher pensions, for an earlier eligibility age, for lower taxes, and for means-testing no longer hold.

---

[17] The differences in the mean rank between a population subgroup and the rest of the population tend to be smaller than the coefficients of the fractional multinomial model because the population subgroup compares a particular socio-demographic variable category (e.g. people aged over 65) with all other people, but the fractional multinomial logit model compares the category with a more tightly defined alternative, e.g. people under 35.

[18] Do these results suggest people have time inconsistent preferences: that they might like low taxes and retirement income when they are young, but higher taxes and retirement incomes pensions when they are eligible for a pension? This is a possibility but we have no way of testing the hypothesis as we only have a single observation for each individual; moreover, the average differences between age groups could reflect cohort effects rather than age effects. The question of time inconsistency is closely tied to whether or not people respond to the survey in a self-interested manner. We discuss this issue further below, and note that while we find greater evidence of self-interested responses than is typically found in these surveys, the effects are rather small.

The results can be restated by highlighting the socio-demographic factors that are associated with each of the survey criteria (i.e. the columns of table 6 and 7). There are three significant results. First, an increase in the age of eligibility is supported by higher income people, those who are more confident about having a comfortable retirement, and people already aged over 65. Secondly, opposition to means-testing is higher amongst higher income people, those who are more confident about having a comfortable retirement, and people already aged over 65 as well. As people with these characteristics are more likely to be of European rather than non-European ethnicity, New Zealanders with European ethnicity are over-represented in the group who support an increase in the age of eligibility and oppose means-testing. Thirdly, saving flexibility is most strongly supported by low income people, those with fewest educational qualifications and – most strongly – by those who are not members of KiwiSaver. The former two groups are the types most likely to need to reduce consumption if there were a compulsory saving scheme, and thus those most likely to be inconvenienced by a compulsory saving scheme.

While these differences in the preferences of different socio-demographic groups are statistically significant, in general they are not large. For example, while universality is more important to people over 65 than under 65, it is the single most important criterion to both groups, and the mean rank for this criteria for those over 65 is only 0.25 less than for those under 65 (2.94 versus 3.20, on a scale of 1 – 7; see Table 6). In the same way, while people over 65 think it is more important to increase the size of the pension than those under 65, the difference between the two groups is only – 0.34 (3.84 versus 4.16, on a scale of 1 - 7). The small size of these effects is in keeping with the international literature, and suggests that socio-demographic characteristics are not the dominant determinants of preferences over retirement income policies. We return to this point in the next section where we divide people into different groups or clusters according to the way they answer the survey and show that similar (but not identical) fractions of each socio-demographic group are in each of the clusters.

While the small size of the differences between different socio-demographic groups is the dominant feature of the results, we find more evidence of self-interested responses than is generally found in the international literature. A feature of the international literature is that there is not much evidence that older people or other socio-demographic groups answer surveys in a particularly self-interested manner, and a lot of evidence that they do not. The argument that survey results only show weak evidence of self-interested behaviour was first made by Ponza et al (1988) and then by Sears and Funk (1990) using U.S. data, and subsequently and forcefully made by Evans and Kelley (2004) using Australian data and Lynch and Myrskylä (2009) using data from 12 European countries. (Hayo and Hiroyuki make a weaker case for Japan). These papers show that in survey after survey the most important determinants of responses on the appropriate structure of government retirement income policy is a respondent's general attitudes, not his or her income or age. To the extent that age or income has a statistically significant effect on survey responses, it is typically very small. Boeri et al (1992) argue there is somewhat stronger evidence of self-interested behaviour in the surveys of Italy and Germany that they consider, but even in these surveys age and income provide little explanatory power of the way individuals respond. Rather, there appears to be considerable diversity of attitudes in all socio-demographic groups.

The results from our survey are more consistent with the hypothesis that people respond in a self-interested manner than the results from these other surveys. As discussed above, we find a small tendency for people in lower income households rather than high

income households to oppose compulsion, to favour means testing, and to favour a lower entitlement age. We also find a small tendency for people over 65 to favour higher pensions, a higher age of eligibility, and to be more opposed to means-testing. It is not clear why these results are more pronounced than those in other surveys, but it may reflect a more accurate way of obtaining preference rankings. Nonetheless, as we have emphasised, the size of these effects is small.

It should be noted that one of the strongest results from the survey does not appear consistent with strong self-interest. 65% of respondents indicated that they would be willing to support an immediate 2 percentage point increase in taxes in order to reduce the size of tax increases on future generations by 2 percentage points, whereas only 30% of respondents indicated the reverse. This result is clearly not consistent with self-interest, even though some of the younger respondents might be expected to benefit from the smaller future tax increases. As we are not aware of similar survey results in the international literature, we cannot compare how this result compares with those from other countries. We should note, however, that it is consistent with the literature arguing that people do not show strong evidence of self-interest in the way that they respond to surveys.

4.4   Cluster Analysis

To investigate the diversity of preferences further, we sorted the respondents into five endogenously determined preference clusters. People in each cluster have reasonably similar preferences: the mean Spearman rank correlations for the members of each cluster ranged from 0.39 to 0.61, much higher than the value of 0.08 estimated for the whole sample. Each cluster contains between 13% and 27% of the respondents. We chose five clusters as the estimated partitions in this case were stable. The results of the cluster analysis are similar if people are allocated to four clusters, as three of the clusters are nearly identical. In contrast, the estimated cluster groups for k = 3 or for k = 6 were not stable.

The five clusters are shown in Table 8. They largely differ by the way their members rank the age of eligibility, universality/means-testing, and saving flexibility/compulsion criteria. With one exception, each of these three criteria are ranked 1 or 2 (most important) or 6 or 7 (least important) in each of the clusters. People in the two largest clusters (cluster 1 – 'Status quo plus compulsion' and cluster 2 'Raise age plus compulsion') concur that universality was the most important criterion and that saving flexibility was the least important criterion, but disagree as to whether increasing the age of eligibility from 65 to 67 was the second most important or the second least important criterion. People in cluster 3 ('Means-test redistribution') favour the introduction of a means-test to fund higher pension payments, and also support compulsion. People in cluster 4 ('Pension minimalists') favour the least government intervention – they are against means-testing and compulsion, and want the age of eligibility increased. The fifth cluster ('No compulsion') is the smallest and least cohesive cluster and comprises people who are unified because they strongly favour saving flexibility rather than compulsion.

Table 8 – The five preference cluster groups

	Cluster 1 Status quo + compulsion	Cluster 2 Raise age + compulsion	Cluster 3 Means-test redistribution	Cluster 4 Pension minimalists	Cluster 5 No compulsion
Central vector	{6243157}	{4652137}	{2451637}	{5743162}	{4253761}
Sample Fraction (s. dev)	27% (0.3%)	24% (0.9%)	18% (0.7%)	17% (0.7%)	14% (0.7%)
Mean Spearman correlation	0.56	0.61	0.48	0.44	0.39
Mean minimum distance	16.6	14.1	20.6	19.0	21.7

Table 9 shows how the population subgroups are allocated across the five clusters. The final column is a multinomial distribution test that the population subgroup and its complement(s) have the same allocation.[19] The difference in the distributions of a population subgroup and the rest of the population is in most cases small, and for most population subgroups it is not possible to reject the hypothesis that the subgroup has the same distribution across the clusters as the overall sample. There are three main exceptions where the allocation across the five clusters is different. First, non-Europeans, particularly Maori, are more likely to be in cluster 5 ('No compulsion') and less likely to be in cluster 2 ('Raise age plus compulsion') than Europeans. Secondly, respondents in high income households are less likely to be in cluster 3 ('Means-tested redistribution') and cluster 5 ('No compulsion') than the general public, although the differences are significant at the five but not one percent level. Thirdly, there are significant differences in the allocation across clusters of the groups that self-identify in terms of their level of confidence that they will have a comfortable retirement. People who do not expect to be comfortable are significantly more likely to be in clusters 3 and 5 ('Means-tested redistribution' and 'No compulsion') and less likely to be in cluster 4 ('Pension minimalists'.) Those who are confident they will be comfortable are much more likely to be in cluster 4 ('Pension minimalists') and much less likely to be in cluster 3 ('Means-tested redistribution'). These are the largest differences between identified population subgroups and reinforce the finding that differences based on non-observable characteristics tend to be larger than differences based on observable characteristics. There are also significant differences in the allocation of KiwiSaver and non-KiwiSaver members across the five groups: non-KiwiSaver members are much more likely to be in cluster 4 ('Pension minimalists') and cluster 5 ('No compulsion') than KiwiSaver members.

In the previous section it was argued that, consistent with the international literature, there was evidence that the responses to the survey reflected self-interest, but that this self-interest effect was small. The literature further suggests that survey responses typically reflect basic philosophical attitudes (such as the relative importance of luck and hard-work in achieving success) rather than identifiable socio-demographic characteristics (Sears and Funk 1990; Fong 2001; Evans and Kelley 2004; Lynch and Myrskylä 2009). There is further evidence that our survey fits this pattern. The difference in the average ranking of each criteria by members of different clusters was 1.52 (on the scale of 1 –7). This difference is several times larger than the difference in the average rankings of different criteria by members of different socio-demographic groups shown in Table 6. The small

---

[19] The asymptotical distribution of the test statistic is χ2 (4).

size of the self-interest effects is reflected in Table 9, which shows that there are only very small differences in the fraction of each identified socio-demographic group in each of the different clusters. Since members of each socio-demographic group are found in similar fractions in each cluster, self-interest effects must be relatively small. Put another way, there are only small differences in the average responses across different socio-demographic groups because the diversity of views within each social-economic group is very similar. This basic finding makes it unlikely that one would find large differences in the preferences of groups identified by other socio-demographic criteria. For example, even though we have no information on whether or not a person is a net tax-payer, the small differences in the allocation of high, middle, and low income groups across the different clusters makes it unlikely that net tax-paying status will be an important determinant of preferences.

Table 9 – Allocation of different population subgroups across clusters

	N	Cluster 1 Status quo compulsion	Cluster 2 Raise age + compulsion	Cluster 3 Means-test redistribute	Cluster 4 Pension minimalists	Cluster 5 No compulsion	Test $\chi^2(4)$
All people		27%	24%	18%	17%	14%	
Demographic characteristics							
Males	510	27%	27%	18%	16%	13%	3.90
Female	556	27%	22%	18%	18%	15%	3.90
Single	253	23%	28%	20%	14%	15%	7.32
Have children	624	27%	22%	18%	18%	16%	7.69
Age <35	308	26%	22%	21%	18%	14%	2.53
Age 35-64	556	29%	23%	17%	17%	15%	3.68
Age 65+	202	24%	32%	17%	16%	12%	8.17
Region and ethnicity							
Auckland	353	27%	28%	16%	15%	15%	7.27
European	783	26%	26%	18%	18%	12%	10.6*
Non-European	283	29%	19%	18%	15%	18%	10.6*
Maori	160	24%	17%	17%	19%	23%	16.6**
Pacific Island	70	27%	23%	22%	6%	20%	7.98
Asian	128	27%	21%	17%	19%	16%	1.39
Highest education, and employment status							
Second school	340	27%	23%	15%	19%	16%	5.96
Degree	466	25%	24%	20%	18%	13%	2.74
Full-time job	558	28%	25%	17%	17%	13%	3.88
Household income							
<$50,000	368	27%	21%	21%	16%	16%	6.90
$50-100,000	426	26%	25%	18%	16%	15%	1.02
>$100,000	272	29%	27%	14%	19%	10%	9.67*

	N	Cluster 1 Status quo compulsion	Cluster 2 Raise age + compulsion	Cluster 3 Means-test redistribute	Cluster 4 Pension minimalists	Cluster 5 No compulsion	Test $\chi^2(4)$
		Confidence about adequacy of retirement income/ KiwiSaver membership					
Not confident	86	19%	17%	30%	8%	26%	25.6**
Very confident	183	26%	27%	9%	25%	13%	19.6**
KiwiSaver	701	29%	26%	19%	15%	11%	21.5**
Non-KiwiSaver	365	23%	21%	16%	21%	19%	21.5**

5   Policy Choices

Preference heterogeneity is a feature of most private markets, and firms develop a wide variety of products to cater to the diverse goods people demand. Preference heterogeneity raises problems for countries that impose mandatory policies on their citizens, however, as policies that are preferred by one population subgroup may be strongly disliked by another. In these circumstances an understanding of the diversity of preferences can be an important input into the policy development process.

This section estimates how the adoption of three fiscally neutral retirement income policies would affect the self-assessed well-being of New Zealanders, given the diversity of their preferences over different retirement income criteria. The three policies have been the subject of extensive discussion in New Zealand. The first policy, 'PAYGO65' is the continuation of New Zealand Superannuation in its current form. It has the following features:

(i)   The weekly payment for single recipients is approximately $370 in 2014/15[20].

(ii)   All people satisfying residency criteria receive the pension when they turn 65.

(iii) The pension is largely funded on a pay-as-you-go basis. No tax increase is required to finance current payments.

(iv) As the population ages, future taxes will have to increase to finance payments. (Projections by the New Zealand Treasury suggest taxes will need to increase by 0.8% of taxable income by 2020 and by 4.6% of contemporaneous taxable income by 2050. We use the latter figure as the proxy for the tax on the next generation.)[21]

(v)   The pension is universal, not means-tested.

The second policy, 'PAYGO67' is similar but the age of eligibility for people born after 1953 is increased to 67 from 2020. No change in current taxes is necessary, but the taxes needed to pay for the pension will only increase by 3.5% of taxable income by 2050. The

---

[20] There are different rates for married and single recipients.

[21] In 2050 many respondents will still be alive and paying tax. Nonetheless, we chose 2050 as an indicative year for the tax rates on future generations because the Treasury routinely produces forecasts of the taxes necessary to pay for expenditure in 2050. The choice of 2050 also allows us to compare the PAYGO65 and SAYGO65 options. The results are robust to the choice of the year.

third policy, 'SAYGO65', keeps the age of eligibility and the structure of pension payments the same as PAYGO65, but a tax surcharge equal to 2% of taxable income is immediately imposed and placed in a sovereign wealth fund, the New Zealand Superannuation Fund.[22] The additional contributions are assumed to cease after 25 years, at which point the ongoing earnings of the fund enable future taxes to be 2% of taxable income lower than they otherwise would be.[23] The essential differences in the three policies are shown in Table 10.

Table 10 – Three possible retirement income policies

	Age of eligibility	Tax increase in 2015	Tax increase in 2050
PAYGO 65	65	0%	4.6%
PAYGO 67	67	0%	3.5%
SAYGO 65	65	2%	2.6%

A key feature of multi-criteria decision making surveys is that the estimated preference rankings can be used to estimate how the welfare of respondents would be affected by a set of policies. When each survey criterion has multiple categories, the willingness of a respondent to trade one aspect of a policy for another can be estimated accurately. Unfortunately, much of this accuracy is lost when there are only two categories for each criterion – it is possible to know, for example, that a respondent would prefer taxes on future generations to increase by 2% rather than to have the age of eligibility increased by two years, but it is not possible to know what level of tax increases would make them indifferent between the two options. Nonetheless, comparisons of policies that differ along a small number of dimensions can still be made. 1000Minds does this by converting the relative ranks of each category into a series of weights that are normalised to sum to one, and then using these weights to calculate a cardinal utility (or well-being) function. The comparisons are made by (i) categorising each policy according to the survey criteria and (ii) estimating the utility of each respondent for each policy using the respondent's own preference weights to make the calculation.

A comparison of the SAYGO65 and the PAYGO65 policies is straightforward to undertake because the effects of the two policies on current and future tax rates differ by exactly the same amount as the difference in the tax criteria categories in the survey. This means each respondent's ranking of the two policies is the same as their relative ranking of the 'current tax' and 'future tax' criteria. The comparison of the PAYGO65 and PAYGO67 policies, and the SAYGO65 and PAYGO67 policies is conceptually more difficult as the difference in the taxes that will need to be imposed on the next generation (either 4.6% and 3.5% of taxable income for PAYGO65 and PAYGO67, or 2.6% and 3.5% of taxable income for SAYGO65 and PAYGO67) is different to the difference in the 'future tax' criterion categories. The 1000Minds algorithm uses linear interpolation to rank policies whose effects are midway between the survey categories: for example, since the two categories for the 'future tax' criterion involve tax increases of either 3% or 5% of taxable, the 4.6% tax increase associated with the PAYGO65 policy averages the 'future tax'

---

[22] The New Zealand Superannuation Fund was established in 2002 for this purpose, and contributions were made until 2008.

[23] We assume 2% of income is added to the New Zealand Superannuation Fund each year and that the Fund compounds at a 4% real rate of return. Income is assumed to increase at 2% per year. After 30 years the accumulated Fund will be 84% of the contemporaneous level of income. The return on this Fund is used to reduce taxes by 2% of income. These figures are necessarily uncertain. If the rate of return is lower than 4% real, the contribution period would have to be increased. As a 4% rate of return is low by historic standards the contribution period could be shorter than 30 years. In any case, a contribution period of 30 years easily meets the criteria of the next generation.

weights corresponding to the 3% and 5% categories by 0.2 and 0.8 respectively. We use this procedure, which is accurate when there are multiple categories in each criterion, but note that since there are only two categories to each criterion there is some loss of accuracy.[24]

Table 11 shows the proportion of the entire sample that has the highest, second highest, and lowest utility scores from each policy option. The table indicates that 58% of respondents would obtain the greatest utility from the SAYGO65 policy, 26% from the current PAYGO65 policy, but only 16% from the PAYGO67 policy. Furthermore, only 16% of the population would have the lowest utility from the SAYGO65 policy, whereas 28% and 56% would have the lowest utility from the PAYGO65 and PAYGO67 policies respectively. These results clearly indicate that, judged in terms of their own preferences, the policy option of raising the age of eligibility is the policy least preferred by most people and most preferred by fewest. In contrast a policy of maintaining the age of eligibility at 65 and prefunding some future New Zealand Superannuation payments is the policy most preferred by the largest number of people and least preferred by the fewest. These results reflect the high rank most respondents place on the importance of avoiding large tax increases on future generations. Some 65% of all respondents indicated they thought it was more important to avoid an increase in taxes on the next generation equal to 2% of taxable income than it was to avoid a similarly sized increase in current taxes, and only 30% indicated the converse (5% were indifferent between the two policies). This preference ranking, which was shared by almost all population subgroups, is one of the strongest findings of the survey.

Table 11 – Welfare ranking of different policies

	Highest well-being	Middle well-being	Lowest well-being
PAYGO65	26%	46%	28%
PAYGO67	16%	28%	56%
SAYGO65	58%	26%	16%

The table shows the fraction of the population giving each policy the 1st, 2nd, or 3rd highest ranking.

The strength of these results changes with some of the underlying assumptions used to parameterise the three policy options. For example, if the 2% tax increase associated with the SAYGO65 policy was only implemented for 15 years, not 25 years, fewer funds would be accumulated in the New Zealand Superannuation Fund and future taxes would only decline by 1% rather than 2% of taxable income from the levels they otherwise would be. Fewer people would support an immediate 2% increase in taxes if it only delivered a 1% future tax reduction. Similarly, more people would be willing to support an increase in the age of eligibility to 67 if it reduced future taxes by a larger amount. Nonetheless, experimentation with these parameters indicates that PAYGO67 would not be a particularly popular policy option even if it reduced future taxes by 2% of taxable income. This is because half of the population thinks it is more important to keep the age at 65

---

[24] Consider the comparison of the PAYGO67 and PAYGO65 polices. PAYGO67 has a two year higher age of eligibility, but future taxes will be 1.1% of taxable income lower under PAYGO67 than PAYGO65. The estimated utility associated with the PAYGO67 policy will be greater than the utility associated with the PAYGO65 policy only if the '3% future tax' policy weight multiplied by 0.55 (=1.1/(5-3)) exceeds the 'age of eligibility' policy weight associated with age 65. In practice this means a respondent will prefer PAYGO67 to PAYGO65 only if he or she has a 'future tax' preference rank that is two or three higher than the preference rank of the 'age of eligibility' criterion e.g. if the 'future tax" criterion is ranked 2 and the 'age of eligibility' criterion is ranked 4 or if 'future tax" is ranked 3 and 'age of eligibility' is ranked 6.

than it is to avoid a 2% increase in current taxes. The importance that a sizeable component of the population places on keeping the age of eligibility at 65 is the basic reason why the PAYGO67 policy maximises well-being for such a small fraction of the population.

Table 12 shows the most preferred and least preferred policy options for different population subgroups. Given the earlier finding that there are only small differences in the preferences of most subgroups, it is not surprising that the various population subgroups rank the policies the same way. The SAYGO65 policy is the most popular and the least unpopular policy option for all the population subgroups analysed, and the PAYGO67 policy is the least popular and most unpopular option. Indeed, the Asian ethnicity group is the only population subgroup in which a majority of people did not rank the SAYGO65 policy as the most preferred, and even for this group it was still the single most preferred policy.

Table 12 – Welfare ranking of different policies by population subgroups

Subgroup	N	Policy that has highest well-being			Policy that has lowest well-being		
		PAYGO65	PAYGO67	SAYGO65	PAYGO65	PAYGO67	SAYGO65
All people	1,066	26%	16%	58%	28%	56%	16%
		Demographic characteristics					
Males	510	27%	17%	56%	29%	54%	17%
Females	556	26%	14%	60%	26%	59%	15%
Single	253	26%	16%	58%	30%	54%	16%
Have children	624	28%	15%	56%	27%	58%	15%
Age <35	308	28%	13%	58%	26%	56%	18%
Age 35-64	556	28%	15%	57%	26%	60%	14%
Age 65+	202	19%	22%	59%	35%	48%	17%
		Region or ethnicity					
Auckland	353	27%	16%	57%	27%	54%	18%
European	783	24%	16%	60%	31%	54%	15%
Maori	160	24%	12%	64%	29%	60%	11%
Pacific Island	70	29%	8%	63%	24%	63%	13%
Asian	128	39%	18%	42%	16%	58%	26%
		Highest education, and employment status					
Second school	340	25%	16%	59%	27%	57%	16%
Degree	466	26%	17%	57%	31%	52%	17%
Full-time job	558	28%	16%	56%	29%	56%	15%

Subgroup	N	Policy that has highest well-being			Policy that has lowest well-being		
		PAYGO65	PAYGO67	SAYGO65	PAYGO65	PAYGO67	SAYGO65
		Household income					
<$50,000	368	30%	14%	56%	24%	61%	16%
$50-100,000	426	23%	15%	61%	27%	58%	15%
>$100,000	272	27%	18%	55%	34%	49%	17%
		Confidence about adequacy of retirement income/ KiwiSaver member					
Not confident	86	27%	15%	58%	35%	51%	14%
Very confident	183	28%	15%	57%	28%	56%	165
KiwiSaver	701	27%	18%	54%	23%	60%	17%
No Kiwisaver	365	24%	16%	60%	29%	54%	17%
		Clusters					
Cluster 1	283	37%	0%	63%	0%	99%	1%
Cluster 2	264	16%	27%	57%	27%	19%	25%
Cluster 3	192	22%	9%	69%	9%	62%	11%
Cluster 4	186	20%	40%	40%	40%	19%	34%
Cluster 5	141	37%	5%	58%	5%	80%	11%

In contrast, the popularity of the policies varies substantially across the preference clusters. This is not surprising, as the clusters endogenously group people who have similar preferences. The PAYGO67 policy option is relatively popular with members of cluster 4 ('Pension minimalists') but with almost no-one else, and it is overwhelmingly the least popular choice with members of clusters 1, 3, and 5. The PAYGO65 policy option is relatively popular amongst clusters 1 ('Status quo plus compulsion') and 5 ('No compulsion') but even for these two cluster groups it was less popular than the SAYGO65 policy.

While we have only formally evaluated three policies as a first-pass demonstration of our approach, it is possible to evaluate many other retirement income policies. For example, we could evaluate whether people would be willing to increase current and future tax rates to increase the size of the pension by $30 per week, or we could analyse the effect of indexing future amounts of New Zealand Superannuation to the consumer price index rather than wages.[25] It is also possible to evaluate how policies that provide people with a range of options might affect the overall welfare of New Zealanders. An example of such a policy is the so-called 'flexible-Superannuation' option that is available in countries such as the United States and the United Kingdom that allows people to defer receiving a pension in exchange for receiving a larger amount. In a New Zealand context, we can evaluate the benefits of providing people with the option of receiving an extra $30 per week if they delay taking a pension until age 67, assuming that this was fiscally neutral.[26] Our survey indicates that this option would be the preferred choice for 46% of

---

[25] However, we could not evaluate a policy that reduced the pension by $30 per week unless we were prepared to assume that the respondents had symmetric preferences over pension increases and decreases. As there is no reason to expect these preferences to be symmetric, it would be necessary to conduct a separate survey with a category explicitly allowing for a decline in the size of the pension to evaluate this option.
[26] We can also evaluate the policy if it were not fiscally neutral, by calculating the changes in taxes necessary to fund it.

respondents, while the remaining 54% would prefer current arrangements; hence a policy that allowed people a choice would be welfare enhancing for 46% of the sample.

6 Conclusions

In recent years, governments around the world have used a variety of methods to improve the ways they develop policies that are better aligned with the wishes of their constituents. At the same time, a goal of the international literature measuring living standards is to improve the way that the welfare implications of different policies are assessed. This paper has investigated whether an approach based on the use of multi-criteria decision making surveys applied to representative samples of the general public can be used to help identify and develop policies that raise well-being. While the specific context has concerned retirement income policies, a goal of the project is to understand the whether the approach is likely to be useful in other policy areas.

The survey proved useful in providing detailed information about the relative importance of specific policy criteria to the population sample. The results suggest that making the pension having universally available and non-means-tested is the most important issue for more than 35% of respondents. There is also a strong preference as to how pensions are funded; in particular respondents prefer to avoid large increases in the tax rates facing future generations, although respondents are not opposed to increases in current taxes if they generate improvements in the structure of pension benefits. There is considerable disagreement about the desirability of raising the age of eligibility from 65 to 67, with equal numbers of people either strongly opposed or unconcerned. Finally, large numbers of people do not appear concerned by the prospect of a compulsory saving scheme. Saving flexibility was the least important issue for more than 40% of respondents, although it was more important to some groups including people living in low-income households, Maori, and those who have chosen not to join KiwiSaver.

A feature of the survey is the way it allows analysis of the diversity of preferences, not just the mean level of preferences. The statistical measures we use show not only that policy preferences about retirement income are diverse, but they depend little on observable characteristics such as age, education, income or ethnicity. Rather, peoples' preferences reflect unobservable characteristics and people can be systematically grouped into clusters who share similar attitudes. Some of these attitudes appear to reflect their expectations about their level of comfort in retirement. However, it likely that these attitudes also reflect deep-seated philosophical approaches to life (Bowles and Gintis 2000). For example, there seem to be an identifiable cluster of people united in a preference for minimalist Government intervention into retirement income policy, and another that is keen on greater redistribution. Future surveys could be designed to further investigate how this diversity of beliefs and attitudes affects the well-being implications of different policy interventions.

A second feature of the approach is that it allows an evaluation of the way respondents rank complex policy option. Studies such as Benjamin et al (2014) attempt to do this by finding out the relative importance of the fundamental factors that determine living standards. This study did not investigate the fundamental factors that people use to assess retirement income policies, choosing instead to get people to rank the importance

of different retirement income policy criteria. The technique allows an assessment of the relative merit of different policies that is based on the way each individual evaluates policies, not just average preferences. The technique is used to make an assessment of three simple retirement income policies: the current policy, a variation in which the age of eligibility is raised by two years to reduce future tax obligations, and another variation in which current taxes are increased to reduce future tax obligations. The results indicate that the policy of raising the age of eligibility by two years maximises well-being for the fewest number of people and lowers it for the largest number. In contrast, a policy that raises current taxes to prevent even larger future tax increases raises the well-being of a majority of the population and, indeed, a majority of almost all population subgroups. Obviously, the decisions to change these policies are political. Nonetheless, just as economic models that show how income is redistributed by different policies provide useful information to policymakers, surveys that provide quantitative information about the distribution of preferences also can provide useful information to policymakers.

In general, we believe the multi-criteria survey approach has considerable potential. Nonetheless, it has limitations that may affect how it is applied in the future. Some of these limitations are reasonably straightforward to correct. For example, because the survey was designed to analyse the relative importance of seven different retirement income criteria, and because the intention was to have a large broadly-based survey, only two category choices or levels were included for each criterion. This reduces the accuracy with which respondents' willingness to trade one criterion for another is estimated. The results that this survey highlights – particularly the bimodal distribution of preferences about the appropriate age of eligibility, and the apparent willingness of people to raise current taxes if this prevents an increase in the taxes paid by future generations – suggest that it would be useful to design subsequent surveys to investigate these three policy criteria more thoroughly. This could be done by dropping some criteria (such as the size of the pension) and adding additional categories to others, or by developing web-based survey tools that allow more in-depth surveys through repeated sampling.

Other issues may be more problematic. The first concerns framing. It is well understood that the way questions are framed can significantly affect the responses that are obtained. Some authors such as Bartels (2003) believe this calls into question whether any survey results can be used. Even if one does not go this far, serious questions should be asked as to how framing affects the results, and it may be appropriate in future research to use multiple versions of the survey, each with different framing. Several questions come to mind in the current survey. For example, in the tax criteria questions the impact of a 2 percentage point tax increase mentioned the weekly tax increase ($20 per week) on an annual salary ($50,000), potentially inducing respondents to think the increase is smaller than it is. As another example, the saving flexibility criteria asked about the effect of a 5% compulsory scheme without discussing the fees that might be involved in such a scheme. These issues may legitimately be a cause for scepticism about the results of any particular survey, whether or not it is conducted using multi-criteria decision making software. If this is the case, the use of several differently frames surveys, possibly applied to smaller samples, may be necessary to build up a convincing body of evidence on an issue.

A second issue concerns the ability of relatively simple surveys to capture the real life complexity of an issue. It may be the case that a proposed policy option involves far more complex outcomes than what survey respondents can be realistically expected to contemplate, or can be realistically expected to answer questions about. Consider, for example, the tax criteria used in the survey. People may well be willing to accept an

increase in the personal tax rate. But does this mean the corporate tax rate should also be increased, and if so can people realistically be expected to have considered the long run implications of a rise in corporate tax rates such as the potential for lower capital accumulation and lower future incomes? If the corporate tax rate is not increased, will this provide tax avoidance opportunities that reduce the integrity of the tax system and raise concerns about fairness? These questions may not be critical for this particular survey, as the tax increases are relatively modest and in any case future tax increases seem inevitable if the retirement income system is not fundamentally changed. Nonetheless, they suggest that the questions that this technique can be used to answer should not be too complex. The implications of different categories of each survey criteria need to be able to be comprehensible to survey respondents, and this may rule out many potential survey topics. Despite these qualifications we believe that this technique will complement existing methods of policy analysis by providing new insights about preferences that can inform policy advice.

A more positive issue concerns cost. As the survey is web-based, it can be 'rolled-out' at low cost once suitable nationally representative web-panels are created. This creates the potential to use differently framed surveys, and also to survey people on different occasions. Moreover, the design of the software means that it is possible to let the general public do the survey at nearly zero additional cost, simply by posting the link on a public website, such as that operated by the Commission for Financial Capability, and letting the survey go "viral". While this process will lose the random selection of the current survey, if the results of the current randomly selected group and the "viral" group are sufficiently similar, it would prove an obvious low cost way of discovering more about the diversity of preferences about different retirement income criteria. Such a trial seems desirable, to ascertain whether multi-criteria decision making software can be used to better understand the diversity of opinion about different policy issues.

References

Arrow, Kenneth J. & Hervé. Raynaud, (1986) *"Social Choice and Multicriterion Decision-Making".* MIT Press.

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.

Bartels, Larry M. (2003) "Is "Popular Rule" possible? Polls, political psychology and democracy". *The Brookings Review* 21(3) 12-15.

Belton, V. & Stewart, T. (2002) "*Multiple criteria decision analysis: an integrated approach".* USA, Kluwer Academic Publishers.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex-Reese Jones (2014) "What do you think would make you happier? What do you think you would choose?". *American Economic Review* 102(5) 2083-2110.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Nichole Szembrot (2014) "Beyond happiness and satisfaction: towards well-being indices based on stated preference". *American Economic Review* 104(9) 2698-2735.

Boeri, Tito, Axel Boersch-Supan, and Guido Tabellini (2002) "Pension Reforms and the opinions of European citizens". *American Economic Review Paper and Proceedings* May 2002 396-401

Bowles, Samuel and Herbert Gintis (2000) "Reciprocity, self-interest, and the welfare state". *The Nordic Journal of Political Economy* 26 (January) 33-53.

Devlin, Nancy J. and Jon Sussex (2011) "Incorporating Multiple Criteria in HTS: methods and processes". (London: Office of Health Economics Research)

Evans, M.D.R. and Jonathon Kelley (2004) "Assessing age pension options: Public Opinion in Australia 1994-2001 with comparisons to Finland and Poland". Melbourne Institute Working Paper 21/04.

Fülöp, J. (2005) "*Introduction to Decision Making methods*". In BDEI-3 Workshop Documents, Washington.

Fong, Christina (2001) "Social preferences, self-interest, and the demand for redistribution". *Journal of Public Economics* 82 225-246.

Fourati, Yosr Abid and Cathal O'Donoghue (2009) "Eliciting individual preferences for pension reform". Institute for the Study of Labour, IZA DP 4479

Gamper, C. D. and C. Turcanu (2007) "On the governmental use of multi-criteria analysis". *Ecological Economics* 62 298-307

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.

Hayo, Bernd and Hiroyuki Ono (2010) " Comparing public attitudes towards providing for the livelihood of the elderly in two aging societies: German and Japan". The *Journal of Socio-Economics* 39 72-80

Jain, Anil K. and Richard C. Dubes (1988) "*Algorithms for Clustering Data".* (Englewood Cliffs, NJ: Prentice Hall.)

Karacaoglu, Girol (2015) "*The New Zealand's Living Standards Framework – A stylised model (Aligning public policy with the ways New Zealanders want to live)*". Forthcoming New Zealand Treasury Working Paper.

Landon, E. L., Jr. (1971) "Order Bias, the Ideal Rating, and the Semantic Differential". *Journal of Marketing Research* 8(3): 375-378.

Lees-Marshment, Jennifer (2015) "*The Ministry of Public Input".* (New York: Palgrave MacMillan)

Lloyd S. P. (1982) "Least squares quantization in PCM". IEEE Transactions on Information Theory 28 129-137.

Lynch, Julia, and Mikko Myrskylä. (2009) "Always the third rail? Pension income and policy preferences in European democracies". *Comparative Political Studies* 42(8) 1068-1097.

Mardin, John I. (1995) *Analyzing and Modeling Rank Data*. London: Chapman and Hall.

Mendoza, G.A. and H Martins (2006) "Multi-criteria decision analysis in natural resource management: A critical review of methods and new modeling paradigms". *Forest Ecology and Management* 230 1-22.

OECD (1998) "Public opinion surveys as an input to administrative reform". Paris: OECD Publishing, SIGMA paper No. 25.

OECD (2001) *Citizens as Partners: OECD Handbook on information, consultation and public participation in policy making* (Paris: OECD).

OECD (2014) How's life in New Zealand? OECD Better Life Initiative.

Ombler, F. & Hansen, P. *1000Minds software*, (2012). Available from www.1000minds.com.

Perreault Jr, W. D. (1975) "Controlling order-effect bias". *Public Opinion Quarterly* 39(4): 544.

Ponza, Michael, Greg J. Duncan, Mary Corcoran, and Fred Groskind. "The guns of autumn? Age differences in support for income transfers to the young and old". *Public Opinion Quarterly* 52, no. 4 (1988): 441-466.

Renn, Ortwin, Thomas Webler, Horst Rakel,Peter Dienel, and Branen Johnson (1993) "Public participation in decision making: a three-step procedure". *Policy Sciences* 26(3) 189-214.

Sears, David O. And Carolyn L. Funk (1990) "The limited effect of economic self-interest on the political attitudes of the mass public". *The Journal of Behavioral Economics* 19(3) 247-271

Soroka, Stuart N. and Christopher Wlezien (2005) "Opinion-policy dynamics: public preferences and public expenditure in the United Kingdom". *British Journal of Political Science* 35 665-689

Van Els, P.J.A., W.A. van den End, and M.C. van Rooij (2003) "Pensions and public opinion: a survey amongst Dutch households". *De Nederlandsche Bank Research Memorandum* 752 2003-18.

Wlezien, Christopher (1995) "The public as thermostat: dynamics of preferences for spending". *American Journal of Political Science* 39(4) 981-1000.