
# Determining benefits-related criteria and weights for prioritising health technologies

Trudy Sullivan

Paul Hansen

Department of Economics

University of Otago

**Citation:** Sullivan, T., Hansen, P. Determining benefits-related criteria and weights for prioritising health technologies. Dunedin: Centre for Health Systems, University of Otago, 2014.

Occasional report 14/01
Published March 2014
Centre for Health Systems
Department of Preventive and Social Medicine
Dunedin School of Medicine
University of Otago
PO Box 913, Dunedin 9054, New Zealand

ISSN: 2253-3109 (Print)
ISSN: 2253-3117 (Online)

This document is available on the University of Otago website
www.otago.ac.nz/healthsystems

# Determining benefits-related criteria and weights for prioritising health technologies

Trudy Sullivan[+] and Paul Hansen[*]

*Department of Economics, University of Otago*

## Abstract

Most approaches to health technology prioritisation based on Multi-Criteria Decision Analysis depend on the determination of appropriate criteria and weights for prioritising technologies. We explain and demonstrate a methodology developed and piloted in New Zealand that involves consulting members of the general population. Six focus groups comprising health care consumers, providers and academics ranked 14 'vignettes' representative of real-world technologies and, stimulated by this ranking exercise, discussed relevant considerations for prioritising technologies. Based on these considerations, we specified six benefits-related criteria for inclusion in a discrete choice experiment administered via an online survey to 322 randomly selected adults. Participants favour technologies that help patients who are, in decreasing order of importance, most in need (mean weight = 0.28), most likely to benefit (0.22), young patients (0.14), patients leading healthy lifestyles (0.13), and patients for whom societal benefits arise from their treatment (0.12) and where no alternative treatment exists (0.11). A cluster analysis of participants' individual weights on these criteria reveals they are not strongly correlated with their demographic and background characteristics. We also demonstrate how these criteria and mean weights can be applied in an imaginary prioritisation exercise.

+*Corresponding author:* Trudy Sullivan, Department of Economics, University of Otago, Dunedin 9054, New Zealand, trudy.sullivan@otago.ac.nz, ph: +64 3 479 4565.

[+]*Acknowledgements*: This paper arose from research for the first author's PhD in the Department of Economics at the University of Otago. She gratefully acknowledges the contribution of her supervisors: Paul Hansen (second author), Paul Thorsnes and Rob Lawson. Thanks also to the health professionals and clinical experts who advised us as we created the health technology vignettes, our friends, family and colleagues who took part in the pilot tests, and the participants in the focus-group meetings and the DCE survey.

# 1. Introduction

All publicly-funded health systems must grapple with how best to allocate their budgets among the myriad health 'technologies' (drugs, devices, equipment, procedures, etc) that are potentially available. This resource-allocation problem – necessitating that new health technologies be prioritised for funding – is likely to intensify due to continuing advances in health technologies, ageing populations and increasing pressure on government finances generally (Appleby, 2013). In the last few years, methods for health technology prioritisation based on Multi-Criteria Decision Analysis (MCDA) have become increasingly popular (Bridges *et al*, 2011; Thokala and Duenas, 2012).

In general terms, MCDA, a sub-discipline of Operations Research, is concerned with decision-making situations in which alternatives are to be ranked based on decision-makers considering a variety of criteria or objectives simultaneously (Belton and Stewart, 2002). Fundamental to MCDA-based approaches to health technology prioritisation is the need to determine appropriate criteria for prioritising technologies and also weights reflecting the criteria's relative importance to decision-makers (Devlin and Sussex, 2011). Previous approaches to determining criteria and weights include public consultation, surveys, focus groups and discrete choice experiments (Casey and Krueger, 2004; Florin and Dixon, 2004; Ryan and Gerard, 2003; de Bekker-Grob *et al*, 2012).

In this paper we explain and demonstrate a methodology for determining criteria and weights that incorporates all of the above-mentioned elements, including focus groups to inform the specification of appropriate criteria for use in a discrete choice experiment (DCE) (McFadden, 1974) to determine the weights on the criteria. Our methodology involves consulting members of the general population, which is justified by the fact that everyone consumes health care during their lifetime and also that most citizens pay taxes and insurance premiums to fund the health system. Therefore, it is appropriate – and a strength of our methodology – that every-day 'ordinary' people determine the criteria and weights for deciding which technologies are available.

Understanding people's preferences is important to prioritisation agencies such as the UK's National Institute of Clinical Excellence and New Zealand's Pharmaceutical Management Agency (PHARMAC). PHARMAC, for example, recently undertook a major public consultation exercise, with the aim of, in the words of Chief Executive Steffan Crausaz, ensuring that "the criteria we use to help us make those decisions … mean our funding decisions continue to reflect the things New Zealanders … value." (PHARMAC, 2013, p. 2). Although our methodology for determining the criteria and weights was developed and piloted in New Zealand, our objective is to explain it such that it could be applied anywhere.

Another strength of our methodology is that the DCE method we use to derive the weights on the criteria yields a full set of weights for each individual participant – in contrast

to most other studies where the DCE method produces aggregated data only (Ryan and Gerard, 2003; de Bekker-Grob *et al*, 2012). These individual-level data allow us to perform cluster analysis (Spath, 1980) to identify 'clusters' of participants with similar patterns of weights.

Our methodology is consistent with the prioritisation framework proposed by Golan and Hansen (2012). In that framework, each technology's incremental benefits – comprising multiple dimensions combined into a single benefits-related variable – is compared with its incremental costs, while also considering the quality of clinical evidence and allowing for, on a technology-by-technology basis, any additional 'X-factors' not elsewhere included, such as strategic or legal factors, etc. It is axiomatic that when thinking about prioritising technologies, most people care about the technologies' costs and quality of clinical evidence. But what do people care about with respect to the multiple dimensions (criteria) underpinning technologies' incremental benefits? And what is the relative importance (weights) of these benefits-related criteria? Our methodology addresses these two questions.

# 2. Methods

## 2.1 Creating health technology 'vignettes'

We began by creating 14 'vignettes' representing real-world technologies to be used in the focus groups at the next step. Each vignette consists of a short description of the technology in terms of clinical indications, treatment and side effects, patient characteristics, numbers treated, etc. Informed by the literature, we assumed that this type of information is pertinent to health technology prioritisation. We exclude technologies' costs and quality of clinical evidence, which, as discussed above, we accept as being axiomatically important, and so these two variables are, in effect, 'put aside' for now; they can be incorporated into the over-arching prioritisation framework later on, as we demonstrate at the end of the paper.

We selected the 14 technologies for the vignettes based on advice from health professionals and personnel from relevant organisations such as PHARMAC. While restricting the number of technologies to a dozen or so – to minimise the responder burden at the next step – the objective was that they be as diverse as possible with respect to the above-mentioned characteristics. Each vignette was written following an extensive review of the medical literature, and checked for accuracy and clarity by a clinical expert in the relevant field, and, where necessary, refined. To ensure a societal perspective, the vignettes are described at the overall technology level across the relevant patient group for the health system as a whole rather than at the individual patient level. To minimise framing effects, they all have the same format with respect to the type and amount of information and their language and lay-out. Finally, the vignettes were pilot-tested with respect to their clarity by having a convenience sample of our friends, family and colleagues read and then rank them

in order of priority (participate in a 'ranking survey', as described below). Where necessary, the vignettes were further refined before being finalised for use in the next step.

## 2.2 Focus groups to identify criteria

With the aim of discovering the considerations that people think are relevant when prioritising health technologies, six focus groups comprising health care consumers, providers and academics were recruited through our professional and personal networks. Ethical approval for the study was granted by the Department of Economics' Ethical Review Body and the Ngai Tāhu Research Consultation Committee at the University of Otago.

Before attending the meeting for their focus group, each person was asked to complete an online 'ranking survey' that involved ranking the 14 health-technology vignettes (as above) with respect to their value to society and hence their relative desirability for being available in the health system. Participants were instructed not to think about the technologies' costs (which, as explained earlier, were not included in the vignettes). The ranking survey was implemented using 1000Minds software (www.1000Minds.com).[1]

Each focus-group meeting followed an identical format, facilitated by the first author (TS), and, with participants' permission, the meeting was recorded and later transcribed. After introductions, the results from the ranking survey – each participant's individual ranking of the technologies and the mean ranking for the group (based on summing each technology's ranks across the participants) – were presented. These results were used to stimulate discussion about the reasoning behind people's rankings of the technologies. As well, the mean ranking was used as the starting point to reach a ranking agreed to by majority consensus. The objective was to discuss the considerations that people think are relevant when prioritising technologies, which as they were teased out were written on a whiteboard by the facilitator. At the end of each meeting, people were asked whether what was written on the whiteboard accurately reflected what had been discussed, and amendments were made if necessary.

## 2.3 Specifying criteria for the discrete choice experiment

The considerations that people think are relevant when prioritising technologies (obtained from the focus groups) need to be specified as benefits-related criteria in a form suitable for the DCE at the next step below. The DCE requires that each criterion has two or more mutually-exclusive levels of 'performance' or severity and that the criteria and their levels are expressed succinctly and in simple language capable of being easily understood by

[1] This software was co-invented by the second author and is freely available for academic and non-commercial use from him or via the above web address.

participants (members of the general population). This requirement for simplicity rules out using Quality-Adjusted Life Years to represent any of the criteria and their levels.

Although, as explained earlier, the vignettes were described at the overall technology level across the relevant patient group for the health system as a whole, the DCE involves the prioritisation of individual patients (hypothetical ones). Accordingly, the criteria for use in the DCE need to be specified in individual patient terms, which also rules out including a criterion associated with 'number of patients'. The design of the DCE in individual patient terms is justified on the grounds that it avoids the following two types of effects capable of biasing DCE responses that have been observed elsewhere.

First, Slovic (1975) and Tversky *et al* (1988) found that many decision-makers, when faced with choosing between alternatives that are, in fact, equally attractive, choose the one that is 'superior' on the criterion that is most important to the decision-maker. This is known as the 'prominence effect'. We suspected that a criterion associated with 'number of patients' might be susceptible to this effect. Moreover, even when a criterion is not truly the most important, it may still be unduly favoured. For decisions that are cognitively challenging – as in DCEs – decision-makers sometimes favour a particular criterion, not because they truly prefer it, but because doing so simplifies and speeds up the decision process. Bryan and Roberts (2008) refers to this phenomenon when it involves favouring health technologies affecting the greatest number of people as the "numbers game nature of a discrete choice approach" (p. 150).

In short, we worried that including 'number of patients' as a criterion in the DCE might encourage some participants to improperly favour technologies with the most patients and to neglect the other criteria. To ensure this did not happen, the DCE is specified in terms of individual patients instead of technologies. The number of patients affected by a technology will be incorporated into the over-arching prioritisation framework later on – along with costs and quality of clinical evidence (as explained earlier).

Another requirement imposed by the DCE is that the levels within each criterion are *a priori* ranked from lowest to highest with respect to their relative importance for determining technologies' priorities. Ideally, these rankings of levels within each criterion would be inherent; for example, everyone would agree that a 'large' health benefit to a patient is better (higher ranked) than a 'medium' benefit, which is better still than a 'small' benefit. If such an inherent ranking is impossible, then a ranking must be imposed – preferably, one that is likely to be as widely accepted as possible.

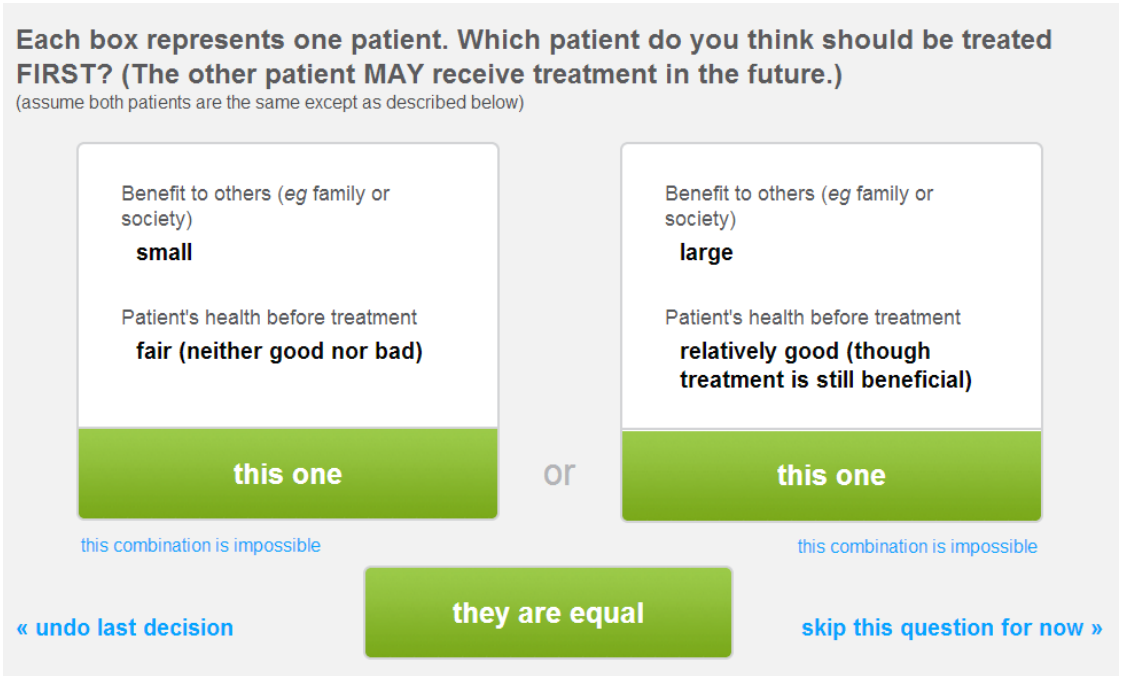
## 2.4 Discrete choice experiment

The final step involves the DCE to determine the weights (often referred to as 'part-worth utilities' in the DCE literature) on the criteria and the levels within each criterion. The DCE was administered via an online survey to which 3218 randomly selected adults drawn from the New Zealand electoral roll were invited to participate.

The online survey was administered using the 1000Minds software referred to earlier in sub-section 2.2. 1000Minds implements the PAPRIKA method (Hansen and Ombler, 2008), a type of 'choice-based' DCE that involves participants pairwise ranking hypothetical patients, as defined on the criteria, with respect to their relative priority for treatment. Other DCE methods and software are also potentially available, as surveyed in Devlin and Sussex (2011), de Bekker-Grob *et al* (2012) and McGinley (2012). Previous research-focussed applications of 1000Minds software in the area of health technology prioritisation include Golan *et al* (2011) and Golan and Hansen (2012). Other 'health' applications include prioritising patients for elective surgery in New Zealand (Taylor and Laking, 2010; Hansen *et al*, 2012) and Canada (Fitzgerald *et al*, 2011), disease classification for rheumatoid arthritis (Aletaha *et al*, 2010) and systemic sclerosis (van den Hoogen *et al*, 2013), measuring clinical trial responses for gout patients (Taylor *et al*, 2011, 2013), and testing physical function for hip or knee replacement patients (Dobson *et al*, 2013).

The 1000Minds software begins by identifying all pairs of hypothetical patients defined on two criteria at a time (in effect, the other criteria are the same) that involve a trade-off. Each participant in the survey is presented with a pair of patients and asked to choose which one should be treated first. An example of a pairwise-ranking question appears in Figure 1.

**Figure 1. Example of a pairwise-ranking question (a screenshot from 1000Minds software)**



The software repeatedly selects pairs of hypothetical patients at random – always involving a trade-off between the two criteria involved – and presents them to the participant for him or her to pairwise rank. Each time the participant ranks a pair, the software immediately identifies all other hypothetical patients that can be pairwise ranked via transitivity and eliminates them from the survey. For example, if a participant prioritises hypothetical patient *A* over patient *B* and then she prioritises *B* over patient *C*, then – by transitivity – *A* is prioritised over *C* (and so the software would not ask a question pertaining to this third pair of patients). This elimination procedure ensures that the number of trade-off questions asked is minimised, the number varying with the choices made. And yet in the process of answering a relatively small number of questions the participant ends up having pairwise ranked all hypothetical patients differentiated on two criteria at a time, either explicitly or implicitly (by transitivity).

Finally, from the participant's explicit pairwise rankings, the software uses mathematical methods based on linear programming to derive weights for the criteria and the levels within each criterion (for technical details, see Hansen and Ombler, 2008). These weights are reported at the individual participant level and also averaged across the group of participants as a whole.

## 2.5 Cluster analysis

As mentioned in the Introduction, a particular strength of the PAPRIKA method relative to other methods for performing a DCE is that a full set of weights is generated for each individual participant, which enables cluster analysis (Spath, 1980) to be performed to identify 'clusters' of participants with similar patterns of weights. As implemented by SPSS statistical software (www.ibm.com/software/analytics/spss), we used Ward's method of clustering (Ward, 1963) with a squared Euclidean measure.

As well as the DCE, survey participants were asked some basic socio-demographic and background questions. These data were used in conjunction with the cluster analysis to examine, via Chi-square tests, the extent to which the identified clusters are correlated with participants' characteristics.

## 2.6 Survey comments and test-retest reliability

Survey participants were also invited to comment on the style of the DCE and the survey overall. In addition, to investigate the reliability of the DCE, a convenience sample of 29 people was recruited to complete the survey twice, approximately 12 days apart. Each individual's criteria weights from both surveys were compared using a paired-samples t-test to determine whether the average difference between the two mean weights for each criterion is statistically significant.

# 3. Results

## 3.1 Health technology vignettes

Vignettes were created for these 14 technologies: abatacept for rheumatoid arthritis, antiretroviral drugs for HIV, dialysis for renal disease, growth hormone treatment, hand sanitiser use in primary schools, hip replacements, imatinib mesylate for chronic myeloid leukaemia, IVF treatment, methadone, oral drugs for erectile dysfunction, PET scan, service for postnatal depression, statins for cardiovascular disease, and vaccine for cervical cancer. The 14 vignettes are presented in the Appendix.

## 3.2 Focus group results

Six focus groups were recruited: Group 1: five general practice staff – a registrar, practice nurse, practice manager and two general practitioners; Group 2: five nurses from the public and private sectors; Group 3: four staff from a non-medical health care organisation; Group 4: six public health professionals and academics – from Public South Health (a regional public health care provider), the Plunket Society (a health care provider for infants and young children) and the Department of Preventive and Social Medicine at the University of Otago; Group 5: 13 staff from a health care provider for Māori (New Zealand's indigenous ethnic minority); and Group 6: seven retirees, all over the age of 65 years.

Each group's majority consensus ranking of the 14 vignettes, as well as mean and median ranks across all groups, are reported in Table I. 'Statins for patients at high risk of cardiovascular disease' is the highest or second-highest priority for all groups. At the other extreme, 'oral drugs for erectile dysfunction' is ranked last or second-last by five groups and third-last by the remaining group. Kendall's coefficient of concordance (W) across the six focus groups is 0.553 (p=0.000) which indicates moderate agreement between the six groups.[2]

The focus-group meetings revealed that people think the following considerations are relevant, at least to some extent, when prioritising technologies (as well as their costs, which participants were instructed to ignore): patients' age, their need and capacity to benefit in terms of length and quality of life and the technology's effectiveness, the extent to which alternative treatments are available, whether or not the illness or injury arose from lifestyle choices, and public benefits including 'equality', 'equity' and 'access' (variously defined). Also, technologies capable of preventing illnesses were generally regarded as superior to curative ones.

[2] If W = 1 then the rankings are identical, whereas if W = 0 then there's no overall agreement between the rankings and they may be regarded as, in essence, random.

**Table I. Rankings of the 14 health technology vignettes by the six focus groups**

| Health technology vignette | Focus group[1] | | | | | | Mean rank[2] | Median rank[3] |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | | |
| Statins for patients at high risk of cardiovascular disease | 1st | 1st | 1st | 1st | 2nd | 2nd | 1.3 | 1 |
| Service for postnatal depression | 6th | 3rd | 7th | 2nd | 1st | 4th | 3.8 | 3.5 |
| Hip replacements | 2nd | 11th | 2nd | 4th | 4th | 3rd | 4.3 | 3.5 |
| Methadone for opioid addiction | 4th | 5th | 6th | 7th | 5th | 10th | 6.2 | 5.5 |
| Vaccine for preventing cervical cancer | 3rd | 13th | 5th | 5th | 3rd | 11th | 6.7 | 5 |
| IVF treatment | 8th | 4th | 8th | 6th | 9th | 8th | 7.2 | 8 |
| Positron emission tomography (PET Scan) | 11th | 12th | 4th | 10th | 6th | 1st | 7.3 | 8 |
| Dialysis for end-stage renal disease | 7th | 7th | 10th | 8th | 7th | 7th | 7.7 | 7 |
| Abatacept for last-line treatment of rheumatoid arthritis | 9th | 6th | 12th | 12th | 8th | 5th | 8.7 | 8.5 |
| Antiretroviral drugs for HIV | 5th | 10th | 9th | 9th | 10th | 9th | 8.7 | 9 |
| Imatinib mesylate for chronic myeloid leukaemia | 10th | 2nd | 11th | 13th | 11th | 6th | 8.8 | 10.5 |
| Hand sanitiser use in primary schools | 12th | 9th | 3rd | 3rd | 12th | 14th | 8.8 | 10.5 |
| Growth hormone for Prader-Willi Syndrome | 13th | 8th | 13th | 14th | 13th | 13th | 12.3 | 13 |
| Oral drugs for erectile dysfunction | 14th | 14th | 14th | 11th | 14th | 12th | 13.2 | 14 |

[1]See sub-section 3.1 for the definitions of the focus groups (1)-(6).
[2]Mean ranks are calculated by summing the group ranks for each vignette and dividing by six, the number of groups.
[3]Median ranks are calculated in the usual way from the group ranks for each vignette.

## 3.3 Criteria for the DCE

Informed by the considerations mentioned above, the six criteria and their levels reported in Table II were specified for the DCE (the weights in Table II are discussed in the next sub-section). An inherent (incontrovertible) ranking of the three age groups 0-14 years, 15-64 and 65+ is impossible, and so, based on Harris's (1985) and Williams' (1997) 'fair innings argument', we ranked the three age groups in the order in the table.

**Table II. Criteria included in the DCE and the mean weights (n=322)**

| Criteria | Mean weights |
|---|---|
| *Patient's health before treatment* | |
| Relatively good (though treatment is still beneficial) | 0 |
| Fair (neither good nor bad) | 0.07 |
| Poor (but not immediately life threatening) | 0.14 |
| Will die soon without treatment | 0.**28**[1] |
| *Benefit to patient (i.e. length and/or quality of life)* | |
| Small | 0 |
| Medium | 0.12 |
| Large | 0.**22** |
| *Age of patient* | |
| 65+ years | 0 |
| 15-64 years | 0.07 |
| 0-14 years | 0.**14** |
| *Illness or injury caused mainly by lifestyle choices* | |
| Yes | 0 |
| No | 0.**13** |
| *Benefit to others (e.g. family or society)* | |
| Small | 0 |
| Large | 0.**12** |
| *Treatment options for this patient* | |
| This is the best treatment (there are less effective alternatives) | 0 |
| This is the **only** treatment available | 0.**11** |

[1]Bolded values represent the relative weights of the criteria overall (i.e. bolded values sum to one).

## 3.4 DCE results

The DCE was completed by 322 participants (a 10% response rate). Their aggregate characteristics are reported in Table III. Compared to New Zealand's adult population as captured in the latest census (Statistics New Zealand, 2013), the sample is fairly representative, though relatively few males, Asian and Pacific Peoples, unqualified and younger people completed the DCE compared to the general population.

**Table III. Average characteristics of DCE participants, and 2013 Census statistics (where comparable)**

| Characteristics | | Percentage (number) | 2013 Census |
|---|---|---|---|
| Gender: | *Male* | 40.4% (130) | 48.5% |
| | *Female* | 59.6% (192) | 51.5% |
| Age (years): | *18-34* | 16.8% (54) | 28.7% |
| | *35-54* | 36% (116) | 32.9% |
| | *55 and over* | 47.2% (152) | 38.4% |
| Ethnicity:[1] | *NZ European* | 85.4% (275) | 71.0% |
| | *Māori* | 8.4% (27) | 11.2% |
| | *Asian* | 3.1% (10) | 11.0% |
| | *Pacific Peoples* | 2.2% (7) | 9.5% |
| | *Other* | 0.9% (3) | 3.0% |
| Income: | *$0-$30,000* | 19.9% (64) | |
| | *$30,001-$70,000* | 31.4% (101) | |
| | *Over $70,000* | 33.5% (108) | |
| | *Not reported* | 15.2% (49) | |
| Qualifications: | *No qualifications* | 2.9% (19) | 20.9% |
| | *Secondary school* | 33.2% (107) | 40.0% |
| | *Other post-secondary school quals* | 25.8% (83) | 19.1% |
| | *University degree or equivalent* | 35.1% (113) | 20.0% |
| Region: | *North Island* | 71.4% (230) | 75.5% |
| | *South Island* | 28.6% (92) | 24.5% |
| Household composition: | *Not living with children* | 60.1% (193) | |
| | *Living with children* | 39.9% (128) | |
| Employment: | *Working* | 62.4% (201) | 62.3% |
| | *Not working* | 15.2% (49) | |
| | *Retired* | 19.6% (63) | |
| Worker type: | *Health-related worker* | 10.2% (33) | |
| | *Non-health-related worker* | 89.8% (289) | |
| Experience of serious illness: | *Yes* | 65.5% (211) | |
| | *No* | 34.5% (111) | |
| Health insurance: | *Yes* | 44.4% (143) | 30.0%[2] |
| | *No* | 55.6% (179) | |
| Health use: | *Seldom* | 18% (58) | |
| | *Occasionally* | 50.6% (163) | |
| | *Frequently* | 30.7% (99) | |

[1]Percentages do not sum to 100, as some people identify with multiple ethnic groups.
[2]Source: Health Funds Association of New Zealand (2013).

To complete the DCE, each participant was required to answer, on average, 25 questions, taking 10-20 minutes in total. The mean weights for the sample are reported in Table II above, where the criteria are listed in decreasing order of relative importance. On average the most important criterion is *patient's health before treatment* (a weight of 0.28) and then *benefit to patient (i.e. length and/or quality of life)* (0.22), followed somewhat distantly by *age of patient* (0.14), *illness or injury caused mainly by lifestyle choices* (0.13), *benefit to others (e.g. family or society* (0.12) and, least importantly, *treatment options for this patient* (0.11).

Thus, in summary, the first two criteria above are relatively important – accounting for half of the overall weight between them (i.e. 0.28 + 0.22 = 0.50) – whereas each of the remaining four criteria, which are approximately equally important, are relatively unimportant – though together they account for half of the overall weight too (0.14 + 0.13 + 0.12 + 0.11 = 0.50).

Another lens through which to view the criterion weights (bolded) in Table II is as ratios, representing the relative importance of the criteria: in other words, their average marginal rates of substitution (Amaya-Amaya *et al*, 2008). These ratios are reported in Table IV.

**Table IV. Relative importance of each criterion (means)**

| | *Treatment options for this patient* | *Age of patient* | *Benefit to others (e.g. family or society)* | *Patient's health before treatment* | *Benefit to patient (i.e. length and/or quality of life)* | *Illness or injury caused mainly by lifestyle choices* |
|---|---|---|---|---|---|---|
| *Treatment options for this patient* | | 0.74 | 0.87 | 0.37 | 0.48 | 0.82 |
| *Age of patient* | 1.35 | | 1.17 | 0.50 | 0.65 | 1.11 |
| *Benefit to others (e.g. family or society)* | 1.15 | 0.85 | | 0.43 | 0.55 | 0.95 |
| *Patient's health before treatment* | 2.07 | 2 | 2.35 | | 1.29 | 2.22 |
| *Benefit to patient (i.e. length and/or quality of life)* | 2.10 | 1.55 | 1.82 | 0.78 | | 1.72 |
| *Illness or injury caused mainly by lifestyle choices* | 1.22 | 0.90 | 1.06 | 0.45 | 0.58 | |

Each number is calculated by dividing the weight corresponding to the highest level of the corresponding row criterion by that of the corresponding column criterion from Table III. For example, the upper-right entry, 0.82 = 10.5/12.8.

## 3.5 Cluster analysis results

Six 'clusters' of participants are identifiable. Their mean criterion weights are reported in Table V. The largest cluster, with 81 members, is Cluster 1, whose mean criteria weights most closely resemble the full sample's weights (Table II). Chi-square tests revealed several minor associations between the clusters and their members' socio-demographic and background characteristics. However, as the statistically significant effects are small and the associations mostly involve only a few participants, there do not appear to be any generalisable patterns that are worthy of being reported.

**Table V. Mean weights on the criteria for the six identified clusters**

| Criteria | Cluster 1 (n=81) | Cluster 2 (n=33) | Cluster 3 (n=59) | Cluster 4 (n=77) | Cluster 5 (n=47) | Cluster 6 (n=25) |
|---|---|---|---|---|---|---|
| Only available treatment | 0.07 | 0.09 | **0.18** | 0.08 | 0.12 | 0.09 |
| Age | 0.11 | 0.17 | 0.11 | 0.14 | 0.13 | **0.30** |
| Societal benefit | 0.14 | **0.20**[1] | 0.11 | 0.13 | 0.06 | 0.06 |
| Need | 0.34 | 0.19 | 0.28 | 0.20 | **0.42** | 0.26 |
| Individual benefit | 0.24 | **0.29** | 0.21 | 0.24 | 0.16 | 0.17 |
| Lifestyle | 0.10 | 0.06 | 0.11 | **0.21** | 0.11 | 0.12 |

[1]Weights in bold identify criteria with noticeably higher weights relative to other clusters.

## 3.6 Survey comments and test-retest reliability

Almost half of the DCE participants included comments in their responses, mostly relating to the criteria and how they answered the DCE questions but also about prioritisation in general.[3] Almost 85% indicated that the survey format was 'easy' or 'very easy' to understand, and almost 60% found the trade-off questions (e.g. Figure 1) difficult to answer.

Finally, for the 29 people who completed the DCE twice, as a check of the survey's reliability, on average 7.4 questions were the same in both surveys. Of these, 4.9 questions were answered identically in both surveys and 2.5 contradictorily (whereby the two surveys' pairwise rankings were opposite, or indifference the first time and strong preference the second). Notwithstanding this apparent inconsistency (an alternative possibility is that some participants changed their minds during the 12 days separating the two surveys), a paired-samples t-test revealed no statistically significant difference between the two surveys' mean weights.

[3] In addition, 281 (87%) requested a copy of the study results – indicating a high degree of engagement in the survey.

# 4. Discussion

Our methodology for determining benefits-related criteria and weights for prioritising health technologies involved four main steps. First, supported by 'expert advice', we created 14 'vignettes' representative of real-world technologies. Next, we asked members of six focus groups to rank the vignettes, with these rankings used to stimulate discussions to discover the considerations that people think are relevant when prioritising technologies. Third, these considerations were specified as benefits-related criteria in a form suitable for use in a DCE involving pairwise ranking (prioritising) hypothetical patients defined on the criteria. Finally, the DCE was administered to a random sample of the population, from which weights representing the relative importance of the criteria to each individual participant and also on average for the group as a whole were derived.

As intended, the 14 vignettes varied considerably with respect to their characteristics that, informed by the prioritisation literature, we assumed would be pertinent to people when prioritising technologies. They ranged from simple preventive interventions (e.g. hand sanitiser) through to life-saving treatments (e.g. drugs for chronic myeloid leukaemia); from treatments involving tens of patients (e.g. last-line treatment for rheumatoid arthritis) through to ones involving tens of thousands (e.g. statins); and also treatments for illnesses caused by lifestyle factors (e.g. methadone, drugs for erectile dysfunction).

The focus-group meetings confirmed that most of the characteristics included in the vignettes were, in fact, pertinent. Interestingly though, especially from a New Zealand perspective, prioritising technologies in favour of Māori per se was not mentioned in the focus groups, even though Māori are over-represented for two of the technologies included in the vignettes (dialysis for end-stage renal disease and statins) and on average Māori have poorer health than non-Māori. Although *age of patient* and *illness or injury caused mainly by lifestyle choices* emerged as relevant in all of the focus groups, in practice it is unlikely that patients would be prioritised according to such criteria as this would be legally regarded as discriminatory in most countries. Nonetheless we retained these criteria for the DCE.

We observed in the focus groups that people found it relatively easy to arrive at consensus rankings for the three or four highest-ranked technologies and likewise for the lowest-ranked ones, but that ranking the 'middle' technologies required a lot more discussion. This observation confirms that *fully* ranking a group of technologies when multiple criteria are involved is difficult, and hence that a more structured decision-making process based on explicit criteria and weights is worthwhile. The six focus groups discussed similar things with respect to relevant considerations for prioritising technologies, suggesting that six focus groups was sufficient for canvassing a wide breadth of views. Because no extra – and, arguably, more subjective – information about the relative importance of these considerations was sought, this methodology based on group discussion proved to be highly effective for merely eliciting them.

As explained in sub-section 2.3, the criteria used in the DCE were specified in individual patient terms instead of at the overall technology level to guard against the prominence effect (Slovic, 1975); Tversky et al, 1988) and the "numbers game nature of a discrete choice approach" (Bryan and Roberts, 2008, p. 150). The focus-group discussions confirmed that when many people benefit from a technology this strongly affects (positively) its relative priority. An obvious drawback of specifying the DCE criteria in individual patient terms is that this precludes a criterion associated with 'number of patients affected'. Given that such a 'scale' variable is relevant for determining a technology's overall affordability when allocating a budget at the aggregate ('meso') level, then, as demonstrated in the next sub-section, this variable needs to be subsequently incorporated into the over-arching prioritisation framework – along with costs and quality of clinical evidence (as mentioned several times already). In contrast, the DCE in Golan and Hansen (2012) was specified at the overall technology level and so it included a 'number of patients affected' criterion in the DCE. Thus, although, as we mentioned in the Introduction, our methodology is consistent with Golan and Hansen's (2012) prioritisation framework it is different in this respect.

We also deliberately excluded technologies' costs from the information contained in both the DCE and the ranking survey (whereby focus-group members ranked the 14 technology vignettes). The cognitive burden in both activities was thus greatly reduced, as participants – mostly members of the general population rather than prioritisation experts – were not required to engage implicitly in cost-effectiveness or cost-benefit analyses. Moreover, as discussed in detail in Bryan et al (2002) and Golan et al (2011), were a 'cost' criterion to be included in the DCE (which it was not), this would likely confuse participants due to uncertainty about how to interpret opportunity costs.

It is reassuring that 85% of DCE participants found the DCE format easy or very easy to follow, and that the DCE questions appeared to exhibit test-retest reliability at least with respect to their ultimate effects on the weights derived from participants' answers. That almost 60% of participants reported finding the trade-off questions difficult to answer is not surprising given that choosing between two hypothetical patients involving a trade-off requires some cognitive effort. However, such pairwise choices are intrinsically cognitively easier than choosing between more than two patients or choosing between patients defined on more than two criteria. Drummond *et al* (2005, p. 145) supports the validity of such methods in general: "The advantage of choice-based methods is that choosing ... is a natural human task at which we all have considerable experience, and furthermore it is observable and verifiable."

The results of the cluster analysis suggest that the variation in participants' preferences, as represented by each participant's set of weights on the criteria, is largely idiosyncratic. Individuals' weights are related more to individuals' personal preferences than to their demographic and background characteristics.

Finally, it is worthwhile reminding ourselves that the ultimate objective of determining criteria and weights is to be able to use them for prioritising health technologies. In the final sub-section, in a similar fashion to the exercise in Golan and Hansen (2012), we demonstrate how the criteria and weights determined in this study can be applied in an imaginary prioritisation exercise involving, for convenience, the 14 technologies for which vignettes were created earlier.

## 4.1 Illustration: Applying the criteria and weights to prioritise technologies

This imaginary prioritisation exercise involves the 14 technologies being rated on the six benefits-related criteria in Table II and then scored using the means weights from the DCE survey also in the table. The first author performed this rating based on her understanding of the technologies – as summarised in the vignettes (see the Appendix) – and, ultimately, her judgment. Bear in mind that this exercise is intended for illustrative purposes only; were it to be done 'for real' a more exacting process based on experts' judgments and 'hard' evidence would be followed. The technologies' ratings and total scores are reported in Table VI, as well as the 'other considerations' of relevance (from the 14 vignettes and other sources): number of patients affected, cost per patient, total cost and quality of clinical evidence.

**Table VI. Ratings on the benefits-related criteria and other considerations for the 14 illustrative technologies**

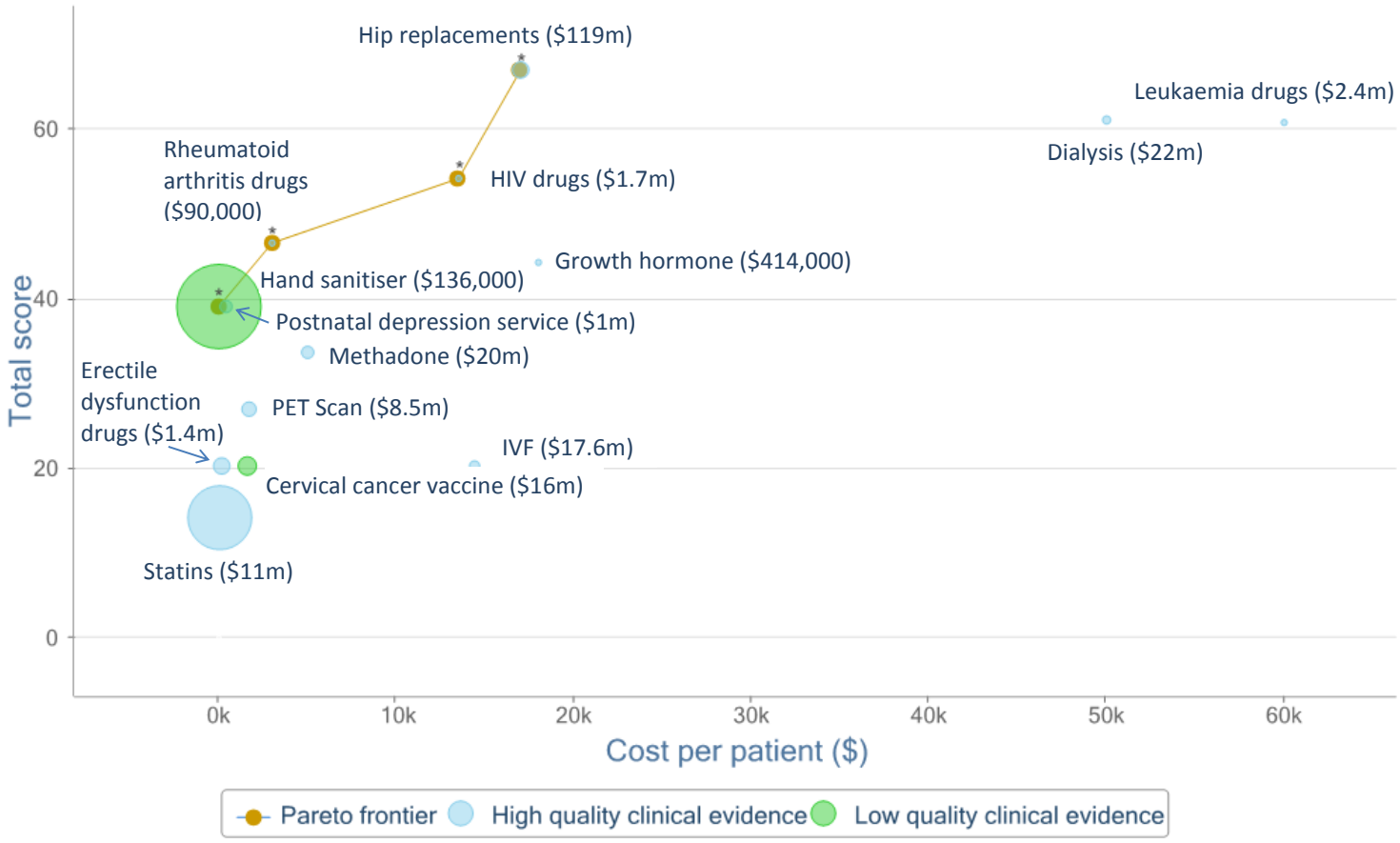
| | Benefits-related criteria | | | | | | | | Other considerations | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Technology | Treatment options | Patient age | Benefit to others (e.g. family or society) | Patient's health before treatment | Benefit to patient (i.e. length and/or quality of life) | Illness or injury caused mainly by lifestyle choices | Total score | Rank | Number of patients affected | Cost per patient ($) | Total cost ($ million) | Quality of clinical evidence |
| Hip replacements | only treatment available | 15-64 | small | poor[1] | large | no | 66.9% | 1st | 7000 | 17,000 | 119 | high |
| Dialysis | only treatment available | 65+ | small | die soon without treatment | large | yes | 61.0% | 2nd | 440 | 50,000 | 22 | high |
| Leukaemia drugs | best treatment available[2] | 15-64 | small | die soon without treatment | medium | no | 60.7% | 3rd | 40 | 60,000 | 2.4 | high |
| HIV drugs | only treatment available | 15-64 | small | poor | large | yes | 54.1% | 4th | 125 | 13,500 | 1.6875 | high |
| Rheumatoid arthritis drugs | best treatment available | 15-64 | small | poor | medium | no | 46.5% | 5th | 30 | 3000 | 0.09 | high |
| Growth hormone | only treatment available | 0-14 | small | fair[3] | small | no | 44.2% | 6th | 23 | 18,000 | 0.414 | high |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hand sanitiser | best treatment available | 0-14 | large | relatively good[4] | small | no | 39.0% | 7th | 400,000 | 0.34 | 0.136 | low |
| Postnatal depression service | best treatment available | 15-64 | large | fair | small | no | 39.0% | 8th | 2500 | 400 | 1 | high |
| Methadone | best treatment available | 15-64 | large | poor | small | yes | 33.6% | 9th | 4000 | 5000 | 20 | high |
| PET Scan | best treatment available | 15-64 | small | fair | small | no | 26.9% | 10th | 5000 | 1700 | 8.5 | high |
| Cervical cancer vaccine | best treatment available | 15-64 | small | relatively good | small | no | 20.2% | 11th= | 10,000 | 1600 | 16 | low |
| Erectile dysfunction drugs | best treatment available | 15-64 | small | relatively good | small | no | 20.2% | 11th= | 9000 | 158 | 1.422 | high |
| IVF | best treatment available | 15-64 | small | relatively good | small | no | 20.2% | 11th= | 1225 | 14,400 | 17.64 | high |
| Statins | best treatment available | 15-64 | small | fair | small | yes | 14.1% | 14th | 220,000 | 50 | 11 | high |

[1] But is not immediately life threatening.
[2] Other less effective alternative treatments are available.
[3]i.e. neither good nor bad.
[4]Treatment is still beneficial.

The main prioritisation variables of interest from Table VI can also be displayed graphically, as in Figure 2. The chart's vertical axis displays each technology's total score, reflecting its aggregate performance on the benefits-related criteria (at the individual patient level), and the horizontal axis displays each technology's cost per patient. The size of the bubble representing each technology is in proportion to the total number of patients affected, and the total cost for the patient group is reported in parenthesises. Finally, the colour of the bubbles indicates the quality of the clinical evidence: blue (or a lighter shade if the chart is in black and white) indicates 'high' quality and 'green' indicates 'low'.

**Figure 2. Main prioritisation variables of interest for the 14 illustrative technologies**




When selecting technologies to be funded, decision-makers should focus their attention first on the technologies in the top-left quadrant of the chart (with high benefits and low cost per patient), while also being mindful of the total number of patients for each technology, the total cost and the quality of clinical evidence. These technologies represent relatively good value for money per patient. In contrast, the technologies in the bottom-right quadrant (low benefits and high cost per patient) represent relatively poor value for money per patient.

The ‘Pareto (efficiency) frontier’ is the line in the chart connecting hand sanitiser, rheumatoid arthritis drugs, HIV drugs and hip replacements; all else being equal, there are no other technologies that have both a lower cost per patient *and* a higher total score (benefit) than these ‘dominant’ technologies. Also relevant is the number of patients, total cost (affordability) and quality of clinical evidence; for example, the effectiveness of hand sanitiser at reducing the spread of germs (compared to using soap and water) is controversial (Stebbins *et al*, 2011), and therefore decision-makers might be reluctant to invest in this technology even at a low cost. By comparing alternative combinations of technologies based on value for money and these other considerations, by a process of ‘trial and error’ decision-makers can arrive at an ‘optimal’ portfolio of technologies (Golan and Hansen, 2012).

# Appendix

The vignettes for the 14 health technologies appear below. Participants in the ranking survey were instructed: “When ranking this treatment, do not consider its cost – just consider its benefits/value to society.”

### 1. Antiretroviral drugs for HIV

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

### 2. Vaccine for preventing cervical cancer (Gardasil)

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

### 3. Dialysis for End-Stage Renal Disease

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

### 4. Growth hormone treatment for Prader-Willi Syndrome

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

### 5. Hip replacements

- A hip replacement is a surgical procedure in which the damaged hip joint is replaced by a prosthetic implant.
- Hip damage is caused by osteoarthritis, rheumatoid arthritis and hip fractures.
- The most common cause of deterioration of the hip joint is osteoarthritis. As the cartilage lining becomes damaged and wears away, the bones within the joint rub together causing pain and making it difficult to get around.

- It can affect men and women, and is more common over the age of 50.
- A hip replacement relieves pain and restores function to the joint. Patients become mobile again and can lead a normal lifestyle.
- A hip replacement typically lasts 15-20 years.
- Number of people to receive a hip replacement: 7000

### 6. Imatinib mesylate for chronic myeloid leukaemia

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

### 7. IVF treatment

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

## 8. Methadone for opioid addiction

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

## 9. Positron Emission Tomography (PET Scan)

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

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

- Erectile dysfunction (ED) occurs when a man is unable to maintain an erection.
- Most ED cases are caused by physical problems, with a small percentage caused by psychological problems.
- Physical problems include high blood pressure, high cholesterol, diabetes, stress, smoking and excessive alcohol intake.
- ED can lead to loss of confidence and self-esteem or depression, as well as to relationship problems.
- Oral drugs such Viagra or Cialis usually allow an erection to occur (with stimulation).

- Around 40% of men over 40 will have erection problems but only 5% will seek help.
- The success rate is approximately 60%.
- Number of men to receive 1 year supply of drugs: 9000 (approximately 2.5% of men with ED).

### 11. Statins for patients at high risk of cardiovascular disease

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

### 12. Abatacept for last-line treatment of rheumatoid arthritis

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

## 13. Hand sanitiser use in primary schools

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

## 14. Service for postnatal depression

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

# References


Aletaha D, Neogi T, Silman A, Funovits J, Felson D, Bingham C, Birnbaum N, Burmester G, Bykerk V, Cohen M, Combe B, Costenbader K, Dougados M, Emery P, Ferraccioli G, Hazes J, Hobbs K, Huizinga T, Kavanaugh A, Kay J, Kvien T, Laing T, Mease P, Ménard H, Moreland L, Naden R, Pincus T, Smolen J, Stanislawska-Biernat E, Symmons D, Tak P, Upchurch K, Vencovský J, Wolfe F, Hawker G. 2010. 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis & Rheumatism* **62**: 2569-2581.

Amaya-Amaya M, Gerard K, Ryan M. 2008. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M (editors). *Using Discrete Choice Experiments to Value Health and Health Care.* Springer: Dordrecht.

Appleby A. 2013. *Spending on Health and Social Care Over the Next 50 Years. Why Think Long Term?* The King's Fund: London.

Belton V, Stewart T. 2002. *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer: Massachusetts.

Bridges J, Hauber A, Marshall D, Lloyd A, Prosser L, Regier D, Johnson F, Mauskopf J. 2011. Conjoint Analysis Applications in Health – A Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health* **14**: 403-413.

Casey M, Krueger R. 2004. An overview of focus group interviewing. In: Roberts AR, Yeager K (editors). *Evidence-based Practice Manual. Research and Outcome Measures in Health and Human Sciences*. Oxford University Press: New York.

de Bekker-Grob E, Ryan M, Gerard K. 2012. Discrete choice experiments in health economics: A review of the literature. *Health Economics* **21**: 145-172.

Devlin N, Sussex J. 2011. *Incorporating Multiple Criteria in HTA: Methods and Processes.* Office of Health Economics, London.

Dobson F, Hinman R, Roos E, Abbott J, Stratford P, Davis A, Buchbinder R, Snyder-Mackler L, Henrotin Y, Thumboo J, Hansen P, Bennell K. 2013. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis & Cartilage* **21**: 1042-1052.

Drummond M, Sculpher M, Torrance GW, O'Brien J, Stoddart G. 2005. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press: Oxford.

Fitzgerald A, De Coster C, McMillan S, Naden R, Armstrong F, Barber A, Cunning L, Conner-Spady B, Hawker G, Lacaille D, Lane C, Mosher D, Rankin J, Sholter D, Noseworthy T. 2011. Relative urgency for referral from primary care to rheumatologists: The Priority Referral Score. *Arthritis Care & Research* **63**: 231-239.

Florin D, Dixon J. 2004. Public involvement in health care. *British Medical Journal* **328**: 159-161.

Golan O, Hansen P, Kaplan G, Tal O. 2011. Health technology prioritization: Which criteria for prioritizing new technologies and what are their relative weights? *Health Policy* **102**: 126-135.

Golan O, Hansen P. 2012. Which health technologies should be funded? A prioritization framework based explicitly on value for money. *Israel Journal of Health Policy Research* **1**: 44.

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**: 87-107.

Hansen P, Hendry A, Naden R, Ombler F, Stewart R. 2012. A new process for creating points systems for prioritizing patients for elective health services. *Clinical Governance: An International Journal* **17**: 200-209.

Harris J. 1985. *The Value of Life*. Routledge: London.

Health Funds Association of New Zealand (2013) Fact File – Health Insurance in New Zealand. www.healthfunds.org.nz/pdf/HFANZ_Fact_File_April_2013.pdf [24 February 2014 ]

McFadden D. 1974. Conditional logit analysis of qualitative choice behaviour. In Zarembka P (editor). *Frontiers in Econometrics*. Academic Press: New York.

McGinley P. 2012. Decision analysis software survey. *OR/MS Today* 39. www.orms-today.org/surveys/das/das.html [24 February 2014]

Ombler F, Hansen P. 1000Minds software, 2012. www.1000minds.com [24 February 2014]

Pharmaceutical Management Agency (PHARMAC). 2013. *PHARMAC's Decision Criteria.* Consultation Documents. PHARMAC: Wellington. www.pharmac.health.nz/ckeditor_assets/attachments/399/consultation-2013-05-17-decision-criteria-review.pdf [24 February 2014]

Ryan M, Gerard K. 2003. Using discrete choice experiments to value health care programmes: Current practice and future research reflections. *Applied Economics and Health Policy* **2**: 55-64.

Spath H. 1980. *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*. Ellis Horwood: New York.

Statistics New Zealand. 2013. *2013 Census QuickStats About National Highlights.* www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-about-national-highlights.aspx [24 February 2014]

Stebbins S, Cummings D, Stark J, Vukotich C, Mitruka K, Thompson W, Rinaldo C, Roth L, Wagner M, Wisniewski S, Dato V, Eng H, Burke D. 2011. Reduction in the incidence of Influenza A but not Influenza B associated with use of hand sanitizer and cough hygiene in schools: A randomized controlled trial. *The Pediatric Infectious Disease Journal* **30**: 921-926.

Taylor W, Laking G. 2010. Value for money – recasting the problem in terms of dynamic access prioritisation. *Disability & Rehabilitation* **32**: 1020.

Taylor W, Singh J, Saag K, Dalbeth N, MacDonald P, Edwards N, Simon L, Stamp L, Neogi T, Gaffo A, Khanna P, Becker M, Schumacher H. 2011. Bringing it all together: A novel approach to the development of response criteria for chronic gout clinical trials. *The Journal of Rheumatology* **38**: 1467-1470.

Taylor W, Brown M, Aati O, Weatherall M, Dalbeth N. 2013. Do patient preferences for core outcome domains for chronic gout studies support the validity of composite response criteria? *Arthritis Care & Research* **65**: 1259-1264.

Thokala P, Duenas A. 2012. Multiple Criteria Decision Analysis for Health Technology Assessment. *Value in Health* **15**: 1172-1181.

van den Hoogen F, Khanna D, Fransen J, Johnson S R, Baron M, Tyndall A, Matucci-Cerinic M, Naden R, Medsger T, Carreira P, Riemekasten G, Clements P, Denton C, Distler O, Allanore Y, Furst D, Gabrielli A, Mayes M, van Laar J, Seibold J, Czirjak L, Steen V, Inanc M, Kowal-Bielecka O, Müller-Ladner U, Valentini G, Veale D, Vonk M, Walker U, Chung L, Collier D, Csuka M, Fessler B, Guiducci S, Herrick A, Hsu V, Jimenez S, Kahaleh B, Merkel P, Sierakowski S, Silver R, Simms R, Varga J, Pope J. 2013. 2013 Classification criteria for systemic sclerosis: An American College of Rheumatology/European League Against Rheumatism Collaborative initiative. *Arthritis & Rheumatism* **65**: 2737-2747.

Ward J. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**: 236-244.

Williams A. 1997. Intergenerational equity: An exploration of the 'Fair Innings' argument. *Health Economics* **6**: 117-132.