THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS

UNIVERSITY OF OTAGO, NEW ZEALAND

The health state preferences and logical inconsistencies of New Zealanders: A tale of two tariffs

N J Devlin
P Hansen
P Kind
A H Williams

DISCUSSION PAPER 180

The health state preferences and logical inconsistencies of New Zealanders: A tale of two tariffs

N J Devlin[1,2], P Hansen[1], P Kind[3], A H Williams[3]

1. *Department of Economics, University of Otago, New Zealand.*
2. *Department of Preventive and Social Medicine, University of Otago, New Zealand.*
3. *Centre for Health Economics, University of York, United Kingdom.*

July 2000

ABSTRACT

Notwithstanding the proposed use of Cost-Utility Analysis (CUA) to inform health care priority setting in New Zealand, to date there has been no research into New Zealanders' valuations of health-related quality of life. This paper reports the results of a study of the health state preferences of adult New Zealanders generated from a postal survey to which 1360 people responded (a 50% response rate). The survey employed a self-completed questionnaire in which a selection of health states were described using the EQ-5D health state classification system and respondents' valuations were sought using a visual analogue scale (VAS). Close attention is paid to the quality of the data, in particular to the 'logical inconsistencies' in respondents' valuations. Regression analysis is used to interpolate values over the 245 possible EQ-5D states. Two tariffs of health state preferences, arising from contrasting treatments of the logical inconsistencies, are reported.

1. INTRODUCTION

Economic analysis is increasingly used to inform decisions about the allocation of resources between alternative health care programmes. In New Zealand particular interest has been shown by key public sector agencies in Cost-Utility Analysis (CUA) to assist in prioritising health spending (HFA 1998, Pharmac 1998) and this has attracted both considerable debate and support (Devlin et al. 1999).

One of the concerns raised is the lack of New Zealand evidence on the valuation of quality of life (Ashton et al. 1999). Notwithstanding the policy relevance of such research, none has been undertaken in New Zealand to date; economic analyses currently rely on health state valuations from other countries for the estimation of Quality-Adjusted Life Years (QALYs). Yet there is the possibility that New Zealanders' health state preferences differ from those of people of other countries. This paper represents a first attempt to explore and report on the health state preferences of New Zealanders. Specifically, our aim is to produce a 'tariff' of health state preferences which will better inform the use of CUA in New Zealand.

A consensus emerging in the literature (Weinstein et al. 1996, Dolan 1999) is that such a tariff is most appropriately based on the values assigned to a set of hypothetical states by the general public (in their dual roles as tax-payers and potential patients) behind 'a veil of ignorance'. This consensus has been accompanied by the development of standardised methodologies for eliciting health state preferences. There are two aspects to this work. The first is the development of generic health state classification systems, which describe 'health' in a general (non-disease-specific) manner, defining it on a number of 'dimensions' such as problems with pain, mobility and so on. Second is the testing of alternative methods of eliciting preferences for health states defined on such systems, including visual analogue scales, standard gamble and time trade-off techniques.

The health state classification system and preference-elicitation method used in the present research is outlined in the following section, followed by a description of the characteristics of the data that were generated from it. Close attention is paid in section 3 to the 'logical inconsistencies' discovered in respondents' valuations. In section 4 the model and regression techniques used to analyse the data are described, and in section 5 the regression results are presented. Section 6 reports the two tariffs of health state preferences that arose from contrasting treatments of the logical inconsistencies in the data. In closing, we draw attention to a number of remaining methodological issues.

2. THE DATA

2.1 The Survey Instrument

This study employs the health state classification system known as the EQ-5D, developed by the EuroQol Group and considered to be one of the systems of choice internationally (Weinstein et al. 1996). Alternative leading systems such as the Health Utilities Index (HUI) have more dimensions and levels with which to represent any given health state, resulting in greater sensitivity but also greater difficulty in eliciting population preferences. The EQ-5D and associated preference elicitation instruments has the additional appeal of being available in the public domain. Given the possibility that this research may have immediate policy relevance, a key consideration has been its timeliness. Consequently, this project is based on data generated from a self-completed questionnaire, in preference to interview-based methods. This in turn dictated the use of a visual analogue scale (VAS) rather than the other techniques mentioned in the Introduction — an issue we will return to in the Discussion.

Figure 1: The EuroQol Health State Classification System, EQ-5D.

Mobility
1. I have no problems walking about.
2. I have some problems walking about.
3. I am confined to bed.

Self-care
1. I have no problems with self-care.
2. I have some problems washing or dressing myself.
3. I am unable to wash or dress myself.

Usual Activities
1. I have no problems with performing my usual activities
(e.g., work, study, leisure or family activities).
2. I have some problems with performing my usual activities.
3. I am unable to perform my usual activities.

Pain/discomfort
1. I have no pain or discomfort.
2. I have moderate pain or discomfort.
3. I have extreme pain or discomfort.

Anxiety/depression
1. I am not anxious or depressed.
2. I am moderately anxious or depressed.
3. I am extremely anxious or depressed.

The EQ-5D describes health states in terms of five dimensions: the degree of mobility, ability to undertake self care, ability to participate in usual activities, degree of pain/discomfort, and degree of pain/depression. Within each dimension there are three levels: (1) no problems, (2) moderate problems, and (3) extreme problems (see Figure 1). Thus each health state can be described by a five-digit combination relating to the relevant level within each dimension (with the dimensions listed in the order given above and in the figure). Together these dimensions and levels define 243 possible health states ranging from 11111 (no problems on any dimension - with value set equal to unity) to 33333 (extreme problems on all dimensions) which, together with ‘unconscious’ and ‘dead’ respectively (the latter set equal to zero), describe 245 states to be valued in a full tariff.

The responder burden imposed by valuation exercises limits the number of states any one respondent can sensibly be asked to value directly. The construction of a tariff is therefore achieved by selecting a sub-set of the 245 states and using these to interpolate values, via regression techniques, for the entire set. To this end, three questionnaires were developed for the study.[1] Version A contained the ‘standard’ set of 13 health states previously used in VAS valuation studies (see Devlin and Williams 1999), in order to facilitate comparative analysis. Versions B and C represented further selections of health states from the array of 245 possible states, a strategy designed to give greater breadth to the observations from which the complete tariff of values is modelled. Table 1 lists the states to be valued in each version of the questionnaire. Respondents were asked to indicate where, on a VAS depicted as a vertical line marked from 100 (“best imaginable health state”) to 0 (”worst imaginable health state”), they considered each state was for them.

Questionnaires were mailed in January 1999, with a reminder and replicate questionnaire two weeks later, to a non-stratified random sample of 3,000 adult New Zealanders on the electoral roll. An initial check confirmed the sample to be representative of the New Zealand population in terms of age, sex and Maori/non-Maori composition. Some 259 questionnaires were returned as the respondent having deceased or no longer living at the address shown on the roll. 1360 completed questionnaires were returned - a response rate of 50% from those who received them.

Table 1: The EQ-5D health states for which valuations were sought.

Version A		Version B		Version C	
11211	21111	11211	11112	11211	11112
11111	11111*	11111	11111*	11111	11111*
21232	unconscious	21111	13311	21111	11113
11122	12111	11133	32211	11133	11312
11121	11112	11121	11131	11121	12111
22233	32211	22222	32313	22222	32223
33333	33333*	33333	33333*	33333	33333*
33321	22323	33323	23232	33323	23232
dead	dead*	dead	dead*	dead	dead*

Note: Health states were displayed over two pages (corresponding to the columns above) in the order listed. States marked * were repeated from the first page.

---

[1] These are available from the authors on request.

2.2 Missing, Implausible and Unusable Data

Of the 1360 completed questionnaires, 103 did not respond to the health state valuation questions, with a further 338 immediately excluded as being unsuitable for estimating a tariff according to the following criteria (see Table 2). If fewer than three states were valued, or if all states were given the same score, they were rejected since this was judged an implausible representation of a respondent's preferences. Also excluded were data that were unusable for the estimation of a tariff due to problems (discussed below) arising from states 11111 and dead.

Table 2: Responses classified according to reasons for exclusion.

Total Responses:	1360
minus *Missing Data*	
Entire data set of valuations missing:	103
Only one or two states valued:	40
minus *Implausible Data*	
All valued states given the same value:	46
minus *Unusable Data*	
Dead and/or 11111 (both observations) not valued:	237
Dead valued greater than or equal to 11111:	15
Useable Responses (i.e., with at least one X_{rescaled}):	919

Note: Some responses suffered from more than one of these defects;
they are classified once only in the order that the criteria appear in the table.

Since health state preferences are anchored for present purposes at dead = 0 and full health = 1, each respondent's valuations (in the range 0 to 100, as above) must be normalised relative to his or her valuations of dead and 11111. Thus the 'raw scores', $X_{\text{raw score}}$, are rescaled according to the transformation $X_{\text{rescaled}} = (X_{\text{raw score}} - \text{dead}_{\text{raw score}})/(11111_{\text{raw score}} - \text{dead}_{\text{raw score}})$. Rescaled values ($X_{\text{rescaled}}$) that fall outside the range 0 to 1 are interpretable as belonging to states considered by the respondent to be worse than dead or better than full health respectively. Given this transformation, if a respondent does not value 11111 and/or dead or scores 11111 $\leq$ dead then X_{rescaled} is either undefined or uninterpretable. Since, as noted above, 11111 and dead are valued twice per questionnaire, all of the respondent's data were excluded if either state is absent from both pages. If absent from one page but not the other, then only the data on the defective page are excluded (these exclusions are not reported in Table 2).

These 919 usable responses constitute the maximum sample from which a tariff of health state preferences can be estimated. The demographic and background characteristics of this 'full' sample are reported in Table 3, as well as, for comparison purposes, the sex, age, ethnicity and smoking characteristics of the New Zealand population (Statistics New Zealand 1997).

Given our aim is to estimate a tariff reflecting the preferences of adult New Zealanders, the sample from which it is estimated ought to be as representative as possible. But, of the characteristics featured in Table 3, which are relevant to explaining variations in individuals' health state valuations? On the basis of Dolan et al.'s (1994) findings for the United Kingdom, of the variables reported in Table 3, education would be expected to exert an influence on health state valuations, while age, sex, smoking, and employment status would not. Comparable research for New Zealand is currently not available, although the data collected for this study will allow us to systematically investigate this question in a future paper.

Nevertheless, one important consideration is New Zealand's unique ethnic makeup, in particular the significant Maori and Pacific Islands minorities. A priori, the ethnicity variable reported in Table 3 is not ruled out as being a potentially significant determinant of New Zealanders' health state valuations. In particular, the possibility that Maori perceive and value health differently than non-Maori is based on what is widely accepted as a 'Maori health perspective', *whare tapa wha*, which compares health to the four walls of a house, each being necessary for symmetry and strength. Each wall represents a different dimension of health: *taha wairua* (the spiritual side), *taha hinengaro* (thoughts and feelings), *taha tinana* (the physical side), and *taha whanau* (the family) (Durie 1998). In this conceptualisation, physical health cannot be distinguished from other dimensions of well-being and the focus of QALYs on effects that are 'health-related' would be seen as an arbitrary distinction. It is also conceivable that Pacific Islands people, New Zealand's other predominant minority ethnic group, perceive and value health differently than Maori and Pakeha/Europeans.

It is clear from Table 3 that Maori and Pacific Islands people are under-represented and Pakeha/Europeans over-represented in the full sample. How serious a short-coming this is impossible to judge; however it should be borne in mind when the tariff from this sample and from an alternative (smaller) sample is presented later in the paper. Similarly, in the light of Dolan's findings regarding the effect of education on health state valuation, the overrepresentation of those with higher education should also be noted. While the form of the education question did not allow direct comparison with Census data (and, conceivably, the sample interpreted the question on 'equivalent professional qualification' rather loosely) it seems clear that lower educated people are under-represented in the sample.

The differences between the full sample and the 441 respondents who were excluded for the reasons noted in Table 2 (i.e., data that were missing, implausible or unusable), although not reported here, are statistically significant. The latter group tended on average to be older (56 years compared to 47), and consequently more likely to be retired. They were also less likely to have continued their schooling after the minimum school-leaving age or to have a degree or equivalent professional qualification, and more likely to be of Maori or Pacific Islands ethnicity. Not surprisingly, given their responses were unusable for our purposes, they were more likely to have reported that they had difficulty completing the questionnaire, although they were no different to the full sample in the extent to which they regarded the descriptions of 'health' employed in the EQ-5D questionnaire as adequate.

Table 3: Demographic and background characteristics (averages) of the full sample, and 1996 Census statistics (where comparable).

	Full sample (n=919) %	1996 Census %
Sex:		
Female	56.1	51.8
Male	43.5	48.2
Age:		
18-19 years	1.7	4.0
20-29	11.6	20.7
30-39	20.9	22.0
40-49	22.4	18.9
50-59	18.4	13.1
60-69	14.8	10.2
70-79	7.1	7.5
80-89	2.3	3.1
90+	0.0	0.4
Ethnicity:		
European or Pakeha	86.1	76.8
Maori	7.6	10.3
Pacific Islands[#]	1.2	3.8
Asian or ‘other’	4.2	4.4
Main Activity:		
Employed/self-employed	62.7	
Retired	17.3	
Houseworker	11.8	
Studying	4.1	
Seeking work	1.6	
Other	2.0	
Did your education continue after the minimum school-leaving age? (yes):	78.3	
Do you have a degree or equivalent professional qualification? (yes):	42.9	
Smoking Behaviour:		
Current smoker	13.3	22.3
Ex-smoker	31.6	21.3
Never-smoker	54.2	48.7
Have you experienced serious illness:		
In yourself? (yes)	28.7	
In your family? (yes)	65.2	
When caring for others? (yes)	37.5	
How did you find filling in this questionnaire?:		
Very difficult	6.6	
Fairly difficult	32.2	
Fairly easy	48.0	
Very easy	11.2	
Do you feel that the way of describing ‘health’ in this questionnaire covers all aspects of your health that are important to you? (yes):	70.7	

Notes: [#] Samoan, Cook Island Maori, Tongan and Niuean. Not all percentages sum to 100 as respondents for whom a response was not recorded are not reported here.

3. LOGICALLY INCONSISTENT DATA

The 919 useable responses can be further classified by the extent to which each exhibits ‘logical inconsistencies’ in the health state valuations, in the following sense. A state with a less severe rating on a particular dimension than another state, given its ratings on the other dimensions are no more severe, can be judged the ‘better’ state and therefore ought, logically, to be scored higher by the respondent making the comparison.[2] For example, 11111 is a better health state than 12111 (and all other health states — except perhaps dead, for religious reasons). Thus a check on the quality of the responses can be made by examining the extent to which such rankings are *not* observed in the data, that is, the number of pairwise ‘logical inconsistencies’ in the reported scores for each respondent. The maximum number of inconsistencies possible is 52 for version A of the questionnaire, 54 for B and 58 for C.[3] Table 4 decomposes the full sample according to the number of inconsistencies exhibited in each useable response.

Table 4: Inconsistencies amongst the full sample (n=919).

Number of pairwise inconsistencies	Number of responses	Cumulative sum of responses
0	189	189
1	207	396
2	137	533
3	98	631
4	56	687
5	48	735
6	30	765
7	18	783
8	17	800
9	14	814
10	12	826
11-20	50	876
21-30	22	898
31-40	13	911
41-50	7	918
57	1	919

Clearly, inconsistencies are the norm. This is not surprising since the questionnaires were self-administered and respondents were not instructed to rank the 14 states before scoring them, as was the case in an interview-based administration of the same instrument for the United Kingdom (Dolan et al. 1994).[4] Also, arguably, for some respondents the ordinal relationships between health states will have been obscured by their description ‘in words’ (as a series of five sentences from Figure 1) rather than numbers (as they are represented in Table 1).

---

[2] Dolan and Kind (1996) describe pairs of valuations that violate this logical principle as ‘primary inconsistencies’ and Badia et al. (1999a) describe them as ‘internal inconsistencies’. Of course, not all pairs can be ranked on this basis, for example, 12111 and 21111.

[3] Note that individuals with missing data are exposed to fewer potential inconsistencies.

[4] However, Dolan and Kind (1996), comparing inconsistencies between self-completed and interview administered VAS questionnaires, concluded that inconsistencies were more common in the latter type, which they suggested to be due to the former being completed by people who were more likely to be able to answer them ‘logically’ — that is, people who were less able to answer logically tended not to answer at all.

The socio-demographic and background characteristics of the individuals with the greatest numbers of inconsistencies were qualitatively similar to the characteristics (not reported) of the 441 respondents who had been immediately excluded from the full sample. In addition, Table 5 presents a comparison of the characteristics for the group with 7 or more inconsistencies (n=154) and the group with 0-1 inconsistencies (n=396). Individuals in the former group tended on average to be older (51 years compared to 45) and were more likely to be retired and to smoke. They were also less likely to have continued schooling after the minimum school-leaving age or to have a degree or equivalent professional qualification, and more likely to be Maori and less likely to be Pakeha/European. Interestingly, they were *less* likely to have reported difficulty completing the questionnaire, which perhaps is explained by their having put less effort into it, or having misunderstood the nature of the task, compared to the group with fewer logical inconsistencies.

Given 80% of the 919 useable responses exhibit logical inconsistencies (Table 4), it must be decided how many of these inconsistencies are acceptable for the purpose of constructing a tariff of health state preferences. That is, what it the appropriate sample, such that we can have confidence in the validity and reliability of respondents' valuations and hence in the tariff derived from them? At one extreme, with as many as 57 inconsistencies, is the full sample (n=919). At the other extreme is a (sub-)sample restricted to individuals with no inconsistencies (n=189); arguably, this is the highest quality data, however it constitutes only a fifth of the otherwise useable responses. Relaxing the number of inconsistencies that are allowed (as follows) permits another six possibilities: 0-1 inconsistencies (n=396), 0-2 inconsistencies (n=533), 0-3 inconsistencies (n=631), 0-4 inconsistencies (n=687), 0-5 inconsistencies (n=735), 0-6 inconsistencies (n=765). Deciding which of these constitutes the appropriate sample from which to estimate the tariff is essentially an arbitrary decision, although some guidance is provided by Ohinmaa and Sintonen (1998).

Using data from a Finnish postal survey similar to ours but without rescaling (this difference is important, as discussed below) they tested the sensitivity of the mean values of health states from different samples, distinguished according to the number of inconsistencies in the individual responses. Compared to individuals with no inconsistencies, the mean values of health states were increasingly different the greater the number of inconsistencies admitted: with more than three inconsistencies, almost 80% of average values were statistically different (at the 5% significance level) from the group with no inconsistencies. Ohinmaa and Sintonen concluded that their estimates were seriously biased by responses with more than three inconsistencies, with the corollary that such responses ought to be excluded. Since "the postal method [for eliciting health state values] will produce significantly more inconsistencies than the VAS studies using interviews"[5] they recommended that "there should be some guidelines [for other researchers] to exclude the most inconsistent respondents from the modeled data set." (p. 15).

---

[5] This statement appears at odds with the conclusion of Dolan and Kind (1996) discussed in footnote 4 above.

Table 5: Demographic and background characteristics (averages) of the respondents with 0-1 and 7+ inconsistencies respectively, and 1996 Census statistics (repeated from Table 3).

	Number of Inconsistencies		
	0-1 (n=396) %	7+ (n=154) %	1996 Census %
Sex:			
Female	54.3	55.2	51.8
Male	45.0	44.8	48.2
Age:			
18-19 years	2.5	2.0	4.0
20-29	14.9***	5.2	20.7
30-39	21.9	19.5	22.0
40-49	22.5	22.1	18.9
50-59	17.4	15.6	13.1
60-69	12.4	22.1**	10.2
70-79	7.1	7.8	7.5
80-89	1.0	5.2**	3.1
90+	0.0	0.0	0.4
Ethnicity:			
European or Pakeha	86.5**	78.4	76.8
Maori	7.1	12.3*	10.3
Pacific Islands[#]	1.4	1.2	3.8
Asian or ‘other’	4.1	6.4	4.4
Main Activity:			
Employed/self-employed	64.1**	52.6	
Retired	15.9	24.7**	
Houseworker	11.6	10.4	
Studying	5.6	3.3	
Seeking work	0.8	3.3*	
Other	2.0	5.8**	
Did your education continue after the minimum school-leaving age?			
(yes):	86.4***	59.7	
Do you have a degree or equivalent professional qualification?			
(yes):	45.5**	34.4	
Smoking Behaviour:			
Current smoker	11.4	19.5**	22.3
Ex-smoker	28.8	31.8	21.3
Never-smoker	58.6**	47.4	48.7
H*ave you experienced serious illness*:			
In yourself? (yes)	23.2	32.5*	
In your family? (yes)	64.9	61.0	
When caring for others? (yes)	34.9	36.4	
How did you find filling in this questionnaire?:			
Very difficult	7.1	9.7	
Fairly difficult	35.4**	24.0	
Fairly easy	45.2	49.4	
Very easy	10.9	13.6	
Do you feel that the way of describing ‘health’ in this questionnaire covers all aspects of your health that are important to you? (yes):	70.5	70.1	

Notes: [#] Samoan, Cook Island Maori, Tongan and Niuean. Not all percentages sum to 100 as respondents for whom a response was not recorded are not reported here. *, **, ***, denotes the proportion is significantly larger than its counterpart in the other sample at the 5%, 2% and 0.2% levels of significance respectively.

No such sensitivity in the mean values of health states is evident in our data. In fact, we find the opposite result: respondents' mean valuations are very robust to the number of inconsistencies admitted. Table 6 reports the mean rescaled values for respondents grouped according to the number of inconsistencies (where an asterisk denotes a statistically significant difference, via a t-test, in the mean value compared to the group with no inconsistencies). As can be seen, almost all of our mean values of X_{rescaled} are statistically the same across groups.

We believe that the absence of an apparent threshold in our data, compared to Ohinmaa and Sintonen's threshold at three inconsistencies, can be explained by our using rescaled values instead of raw scores, as they did.[6] Given that rescaled data are the appropriate data for estimating a tariff, and given that (explained below) t-values are *not* invariant to the data transformations introduced by rescaling (or not), then Ohinmaa and Sintonen's t-tests and the inferences they derive are invalid.

The difference between t-tests based on rescaled data (as reported here) and t-tests from raw data (as reported by Ohinmaa and Sintonen) can be appreciated by considering the usual t-statistic formula, $t = (\overline{X}_0 - \overline{X}_a) / \sqrt{s_0^2 / n_0 + s_a^2 / n_0}$, where $\overline{X}_0$ and $\overline{X}_a$ are the mean health state values for the group with no inconsistencies and its comparator group respectively, s_0^2 and s_a^2 are their variances and n_0 and n_a their sample sizes. Both $\overline{X}_0$ and $\overline{X}_a$ depend (by design) on whether the health state values from which they are calculated are rescaled or not, as do s_0^2 and s_a^2, but the transformation alters their ratio in the t-statistic. In other words, the raw data and the rescaled data produce different t-statistics and hence different inferences.[7] Hence, given that rescaling is necessary for estimating a tariff, Ohinmaa and Sintonen's discovery of an inconsistency threshold, based as it is on raw data only, is invalid, whereas our finding is valid that a threshold does not exist.

This lends prima facie support for the full sample (n=919) being the 'default' for estimating a tariff.[8] The full sample also has the appeal of having arisen 'naturally', rather than by researcher decree as to what is and what is not an acceptable number of inconsistencies in a respondent's valuations. However, two other characteristics of the mean values reported in Table 6 suggest

---

[6] Ohinmaa and Sintonen did this because they wanted to maximise the size of their data set by including respondents who did not value dead which, if they rescaled, would have had to be excluded (as ours were).

[7] Indeed, we believe that using raw data systematically biases the results in favour of finding significant mean value differences. This is because the estimated variances *relative* to the means (this relationship is the basis of the t-test) are likely to be smaller using raw data than if rescaled data are used. This property of the variances can be inferred heuristically (somewhat unsatisfactorily, we concede) by considering the sample ranges. Without scaling the maximum possible range that a health state value can have is 100 – 0 = 100, with the mean between these extremes. With scaling this range has the potential to *increase* to 100 – -99 = 199. That is, at one extreme, if the respondent (raw) scores dead = 0, 11111 = 1 and the state to be rescaled is scored 100, then the rescaled value is 100. At the other extreme, if dead = 99, 11111 = 100 and the state to be rescaled is 0, then the rescaled value is -99. If we are able to infer from this heuristic illustration that the variance relative to the means will be smaller using raw data, compared to rescaled data, then there is a greater likelihood of rejecting the null hypothesis that the sample means are the same, since the calculated t-value will be relatively large.

[8] This would be consistent with the principle agreed to at the Barcelona meeting of the EQ-net project in January 1999 that the number of exclusions should be minimised.

that, in fact, there are systematic differences between the groups such that not all of them belong together in a single sample.

One, the rankings (shown in parentheses in Table 6) increasingly disagree as the number of inconsistencies increase. Two, the number of inconsistencies in the mean X_{rescaled} values vary (in general, increase) across the groups (where, in the table, inconsistent states have a box around their mean value): zero for the group with no (individual) inconsistencies; zero for the group with one inconsistency; two for the group with two inconsistencies;[9] three for the group with three;[10] six for the group with four;[11] three for the group with five;[12] four for the group with six;[13] and 14 for the group with seven-or-more inconsistencies.

Table 6: Mean rescaled health state values (and rankings) grouped according to the number of inconsistencies in the individual respondent scores.

	Mean values of X_{rescaled}														
	Groups, according to the number of inconsistencies in the individual responses														
Health state	0	1		2		3		4		5		6		7+	
11211	0.7931	0.7991	(1)	0.7955	(1)	0.8675	(1)	0.8059	(1)	0.7790	(1)	0.8536	(1)	0.9688	(1)
21111	0.7588	0.7280*	(2)	0.6972*	(4)	0.7370	(3)	0.6239	(5)	0.6446*	(4)	0.7874	(2)	0.6915	(6)
11112	0.7514	0.7097*	(4)	0.7106	(3)	0.7207	(4)	0.7064	(2)	0.7405	(2)	0.7309	(6)	0.9828	(2)
11121	0.7441	0.7274	(3)	0.7186	(2)	0.7846	(2)	0.6733	(4)	0.7310	(3)	0.7838	(3)	0.8467	(3)
12111	0.7057	0.6871	(5)	0.6624	(5)	0.6553	(5)	0.6978	(3)	0.6120	(5)	0.6748	(5)	0.5306	(9)
11131	0.5244	0.4729	(7)	0.4534	(6)	$\boxed{0.4467}$	(8)	0.3506*	(10)	0.5300	(7)	$\boxed{0.3852}$	(9)	0.8241	(4)
11122	0.5203	0.5015	(6)	0.3478	(9)	0.6151	(6)	0.4827	(6)	0.5552	(6)	0.5502	(6)	0.6762	(7)
13311	0.4557	0.4275	(8)	0.3926	(8)	0.3113*	(9)	0.3579	(9)	0.3424	(10)	0.3460	(10)	$\boxed{0.5855}$	(8)
11113	0.4121	0.4200	(9)	0.3979	(7)	0.3043	(11)	0.3864	(8)	0.3538	(9)	0.4409	(7)	$\boxed{0.5167}$	(10)
22222	0.3752	0.3637	(11)	0.3293	(11)	0.3057	(10)	0.3418	(11)	$\boxed{0.2506}$	(14)	0.2860	(12)	$\boxed{0.3992}$	(13)
11312	0.3636	0.3744	(10)	0.3464	(10)	0.2767	(12)	0.4173	(7)	0.2604	(13)	0.2854	(13)	$\boxed{0.4176}$	(12)
32211	0.3360	0.2595*	(12)	0.2143	(12)	0.2653	(13)	$\boxed{0.1608^*}$	(15)	0.2918	(12)	0.2978	(11)	$\boxed{0.0104}$	(22)
21232	0.2923	0.2402	(13)	0.1191	(15)	$\boxed{0.4510^*}$	(7)	0.3159	(12)	0.4910	(8)	$\boxed{0.4333}$	(8)	0.4494	(11)
32313	0.1716	0.1252	(17)	0.1418	(13)	0.0640	(18)	$\boxed{-0.0315^*}$	(21)	0.0599	(17)	0.1499	(14)	$\boxed{0.1093}$	(19)
11133	0.1667	0.1538	(14)	0.1327	(14)	$\boxed{0.0130}$	(19)	$\boxed{0.1555}$	(16)	$\boxed{0.0608}$	(16)	$\boxed{0.0858}$	(19)	$\boxed{0.1514}$	(17)
22323	0.1656	0.1253	(16)	$\boxed{-0.0385}$	(20)	0.2365	(14)	$\boxed{-0.0434}$	(22)	0.0175	(18)	0.0900	(18)	$\boxed{0.7706}$	(5)
22233	0.1506	0.1362	(15)	-0.0080	(19)	$\boxed{0.2160}$	(15)	$\boxed{0.1940}$	(14)	$\boxed{0.2953}$	(11)	$\boxed{0.1258}$	(15)	$\boxed{0.2313}$	(16)
32223	0.1435	0.1061	(19)	0.1167	(16)	-0.0379	(21)	$\boxed{0.2178}$	(13)	-0.0297*	(21)	0.1076	(16)	$\boxed{0.2565}$	(14)
33321	0.1320	0.1045	(20)	$\boxed{-0.0404}$	(21)	0.0950	(16)	0.0721	(17)	0.1919	(15)	0.0693	(20)	$\boxed{-0.0854}$	(24)
23232	0.1122	0.1165	(18)	0.1034	(17)	0.0097	(20)	0.0633	(19)	-0.0009	(20)	$\boxed{0.0194}$	(22)	$\boxed{0.0567}$	(21)
uncon.	0.0326	0.0308	(21)	-0.0862	(23)	0.0728	(17)	-0.1645	(24)	0.0132	(19)	0.1034	(17)	-0.0540	(23)
33323	0.0280	-0.0011	(22)	$\boxed{0.0081}$	(18)	$\boxed{-0.1230}$	(23)	$\boxed{0.0711}$	(18)	$\boxed{-0.0872}$	(23)	$\boxed{-0.0255}$	(24)	$\boxed{0.2393}$	(15)
33333b	-0.0622	-0.0835	(24)	-0.0940	(24)	$\boxed{-0.1108}$	(22)	-0.0634	(23)	$\boxed{-0.0585}$	(22)	-0.0274	(23)	$\boxed{0.1246}$	(18)
33333a	-0.0635	-0.0856	(23)	-0.0854	(22)	-0.1582	(24)	$\boxed{0.0074}$	(20)	-0.0971	(24)	$\boxed{0.0447}$	(21)	$\boxed{0.0958}$	(20)

Notes: Figures in parentheses are rank orders and * denotes that the difference in the mean value, compared to the group with no inconsistencies, is statistically significant at the 5% level. 33333a and 33333b distinguish the two valuations for this state (i.e., from the two page referred to in Table 1). Boxes around values signify inconsistent states (see footnotes 9-13).

---

[9] 33323 is valued above 33321 and 22323.

[10] 33333b above 33323, 22233 above 11133 and 21232 above 11131.

[11] 33323 and 33333a above 22323 and 32313, 22233 above 11133 and 32223 above 32211.

[12] 33333b above 33323 and 22233 above 11133 and 22222.

[13] 33333a above 33323 and 23232, 22233 above 11133 and 21232 above 11131.

These results imply a degree of *systematic*, as opposed to purely idiosyncratic, inconsistency in the groups (other than the ones with zero and one inconsistency respectively) since were they purely idiosyncratic in nature they would be expected to 'cancel each other out' when combined (i.e., averaged) in the groups. Thus the (sub-)sample with 0-1 inconsistencies (n=396) is the leading candidate as an alternative (and comparator) to the full sample (n=919). Bearing in mind our earlier discussion of the potential importance of ethnic differences between respondents, it turns out too that this sample, of the seven possibilities noted earlier, is most representative of the ethnic composition of the New Zealand population (while still under-representing Maori and Pacific Islands people). Furthermore, since this sub-sample constitutes the opposite extreme to the admission of *all* inconsistencies, as in the full sample, it permits us to explore the differences between the two tariffs that are produced from each. We proceed, therefore, in the remainder of the paper, to estimate a tariff of health state preferences from the full sample (n=919) and another from the sub-sample with 0-1 inconsistencies (n=396).

4. STATISTICAL METHODS

Our approach is based on the same model and generalised least-squares regression technique as Dolan (1997). These two aspects are discussed in turn.

4.1 The Model

Each of the 22 health states for which we have data (see Table 6), but not including 'unconscious', can be represented by dummy variables designed to capture possible independent and interaction effects across the five dimensions of the EQ-5D health state classification system. Also, each dimension can be represented as either linear or non-linear in its three levels. Figure 2 (adapted from Dolan, pp. 1099-1100) lists the available dummies, which are grouped in five sets. Like Dolan, we do not have a theoretical model to direct how these dummies ought to be combined in an equation suitable for estimation. Nonetheless we considered it better to include *all* or *none* of the dummies from a particular set in the equations we experimented with, rather than 'mix-and-match' from different sets.

Since the EQ-5D system values health states other than full health as negative deviations from a value of unity (since 11111 = 1), X_{rescaled} can be represented as *1 minus* the appropriate (linear) combination of dummy variables and their coefficients (to be estimated).[14] However, a constant of unity corresponds to transforming the dependent variable to $1 – X_{\text{rescaled}}$ (of which, accordingly, higher values correspond to more severe health states). Thus we experimented with the nine

---

[14] For an illustration of this calculation readers are invited to skip forward to Table 7.

specifications below of the general form:[15] $1–X_{\text{rescaled}}$ = *constant* + $\boldsymbol{D\beta}$ + *error term* (where $\boldsymbol{D}$ is the row vector of dummies (as below) and β the column vector of coefficients). Although transforming the dependent variable to $1–X_{\text{rescaled}}$ implies *constant* = 0, this restriction was not imposed - since were it invalid the estimates of β would be biased (and, possibly, the R^2 statistics outside their usual range of 0 to 1). (As happens, however, the *constant* estimates *are* significantly different from zero - the interpretation of which we consider in the Results below.)

Specification (1): (*MO*, *SC*, *UA*, *PD*, *AD*) - i.e., set 1.
Specification (2): (*MO*, *SC*, *UA*, *PD*, *AD*) and *N3* - set 1 plus *N3*.
Specification (3): (*MO*, *SC*, *UA*, *PD*, *AD*) and (*M2*, *S2*, *U2*, *P2*, *A2*) - sets 1 and 2.
Specification (4): (*MO*, *SC*, *UA*, *PD*, *AD*), (*M2*, *S2*, *U2*, *P2*, *A2*) and *N3* - sets 1 and 2 plus *N3*.
Specification (5): (*MO*, *SC*, *UA*, *PD*, *AD*), (*M2*, *S2*, *U2*, *P2*, *A2*) and (*MOSC*, *MOUA*, *MOPD*, *MOAD*, *SCUA*, *SCPD*, *SCAD*, *UAPD*, *UAAD*, *PDAD*) - sets 1, 2 and 3.
Specification (6): (*MO*, *SC*, *UA*, *PD*, *AD*) and (*MOSC*, *MOUA*, *MOPD*, *MOAD*, *SCUA*, *SCPD*, *SCAD*, *UAPD*, *UAAD*, *PDAD*) - sets 1 and 3.
Specification (7): (*F11*, *F21*, *F31*, *F41*) and (*F13*, *F23*, *F33*, *F43*, *F53*) - sets 4 and 5.
Specification (8): (*MO*, *SC*, *UA*, *PD*, *AD*), (*F11*, *F21*, *F31*, *F41*) and (*F13*, *F23*, *F33*, *F43*, *F53*) - sets 1, 4 and 5.
Specification (9): (*MO*, *SC*, *UA*, *PD*, *AD*) and (*F13*, *F23*, *F33*, *F43*, *F53*) - sets 1 and 5.

Specification (1) treats movements on a particular dimension between levels 1 and 2 and levels 2 and 3 respectively as having identical effects on health state valuations, while (2) recognises an additional effect, captured by the *N3* variable, arising from any of the dimensions being at level 3 (extreme problems). Specifications (3), (4) and (5) allow for non-linearities between levels, with (4), the so-called "main effects" models augmented with the N3 variable, the basis of the equation from which Dolan estimated his tariff. As well, specification (5) allows for interaction effects between dimensions, as does (6). Specification (7) consists entirely of dummies that simply count the number of dimensions at levels 1 and 3 respectively, without distinguishing between dimensions (e.g., 31111 and 11113 are equivalent). As such it is not a serious contender for estimating a tariff since a priori we expect the dimensions to have unique effects on preference values. Specifications (8) and (9) overcome this potential shortcoming by augmenting the dummies in (7) with others that distinguish between particular dimensions.

Our principal criterion for discriminating amongst the estimates of the nine specifications is that the tariff of health state preferences that is produced from it be logically consistent. This imposes several restrictions on the specifications' coefficient point estimates that are discussed in the Results.

---

[15] Other obvious combinations, such as (*M2*, *S2*, *U2*, *P2*, *A2*) with (*F13*, *F23*, *F33*, *F43*, *F53*), are ruled out because they are collinear.

Figure 2: Dummy variables available to model dependent variable $1–X_{\text{rescaled}}$

Set 1: Dummies to represent the (assumed equal) move between all three levels.

MO	1 if mobility is level 2; 2 if mobility is level 3; 0 otherwise
SC	1 if self-care is level 2; 2 if self-care is level 3; 0 otherwise
UA	1 if usual activities is level 2; 2 if usual activities is level 3; 0 otherwise
PD	1 if pain/discomfort is level 2; 2 if pain/discomfort is level 3; 0 otherwise
AD	1 if anxiety/depression is level 2; 2 if anxiety/depression is level 3; 0 otherwise

Set 2: Dummies to represent the move from level 2 to level 3. (This allows the effect of the move from level 1 to level 2 to be different from the effect of the move from level 2 to 3.)

M2	1 if mobility is level 3; 0 otherwise
S2	1 if self-care is level 3; 0 otherwise
U2	1 if usual activities is level 3; 0 otherwise
P2	1 if pain/discomfort is level 3; 0 otherwise
A2	1 if anxiety/depression is level 3; 0 otherwise

Set 3: Dummies to allow for possible (first order) interactions between dimensions.

MOSC	The product of *MO* and *SC*
MOUA	The product of *MO* and *UA*
MOPD	The product of *MO* and *PD*
MOAD	The product of *MO* and *AD*
SCUA	The product of *SC* and *UA*
SCPD	The product of *SC* and *PD*
SCAD	The product of *SC* and *AD*
UAPD	The product of *UA* and *PD*
UAAD	The product of *UA* and *AD*
PDAD	The product of *PD* and *AD*

Set 4: Dummies to count for the number of dimension(s) that are at level 1.

F11	1 if the health state contains 1 dimension at level 1; 0 otherwise
F21	1 if the health state contains 2 dimensions at level 1; 0 otherwise
F31	1 if the health state contains 3 dimensions at level 1; 0 otherwise
F41	1 if the health state contains 4 dimensions at level 1; 0 otherwise

Set 5: Dummies to count the number of dimension(s) that are at level 3.

F13	1 if the health state contains 1 dimension at level 3; 0 otherwise
F23	1 if the health state contains 2 dimensions at level 3; 0 otherwise
F33	1 if the health state contains 3 dimensions at level 3; 0 otherwise
F43	1 if the health state contains 4 dimensions at level 3; 0 otherwise
F53	1 if the health state contains 5 dimensions at level 3; 0 otherwise

Also, in keeping with Dolan: “one further dummy that represents whether any of the dimensions are at level 3.”

N3	1 if any dimension is level 3; 0 otherwise

4.2 Regression Technique

The nine specifications outlined above were estimated as ‘random effects’ models using the LIMDEP programme, version 6 (Greene 1991). The defining feature of this model (in its ‘one factor’ version) is its error term has two parts: a traditional error that is unique to each observation of the dependent variable ($1–X_{\text{rescaled}}$) — irrespective of the individual in the sample whom it concerns — and another that is peculiar to each individual — that represents the extent to which his/her intercept differs from the overall intercept of the model. The purpose of the second component is to allow individual-specific effects for explaining the variation in the dependent variable without having to resort to a dummy variable for each individual (the latter is a ‘fixed effects’ model — e.g., see Greene 1991, pp. 298-9).[16] Thus a given individual who, because of background or idiosyncratic factors, has a tendency to, say, under-value health states relative to other individuals’ valuations will, due to the second error component, have *his/her* estimated intercept value increased without affecting the intercepts of the other sample members.

5. REGRESSION RESULTS

The estimates of the nine specifications for the full sample (n=919) and the sub-sample with 0-1 inconsistencies (n=396) are reported in Tables A1 and A2 in the Appendix. The R^2s for the latter sample are all around 70%, which is a very good fit. For the full sample they are less impressive (around 9%), which perhaps is not surprising given the potentially disparate nature of the data set as a result of the full complement of inconsistencies being admitted.

Notwithstanding very different fits over the two samples, the signs on the estimated coefficients are in general agreement. Moreover, in all six of the alternative sub-samples that we rejected in the Data section (i.e., 0, 0-2, 0-3, 0-4, 0-5 and 0-6 inconsistencies respectively), the signs and significance of the coefficients were mostly replicated.[17] This confirmed our earlier conclusion that, empirically, a threshold in terms of the number of inconsistencies does not exist. As before, this is in contrast to Ohinmaa and Sintonen (1998, p. 12), who, upon estimating specification (3) (recall, with unscaled data) found three inconsistent states to be the maximum number of inconsistencies to not significantly affect the estimates.

As noted in the previous section, choosing between the nine estimated equations hinges on the logical consistency of the tariffs that are generated from their point estimates. For the n=919 sample, only specification (1) satisfies this criterion. For the n=396 sample, specifications (1), (2) and (7) satisfy it; however (2) is the preferred equation since, as discussed earlier, (7) was not a serious contender for estimating a tariff and (1) is encompassed by (2). Furthermore, all of the

---

[16] Given that many of the responses in our data set are incomplete, that is, not all health states in the questionnaire were scored, an advantage of using LIMDEP is that each individual is not required to have the same number of observations.

[17] With R^2s ranging from 78% to 40%. These estimation results are available from the authors on request.

constants are statistically significant (despite, as noted in the previous section, the theory suggesting they ought *not* to be).

How do the logical inconsistencies in the rejected specifications manifest themselves in the estimates reported in Tables A1 and A2? With respect to the full sample, specification (2) is disqualified for estimating a tariff because the insignificant *UA* coefficient dictates that two health profiles with levels 1 and 2 on this dimension respectively, and levels in common for the other four dimensions, are valued the same.[18] Specification (3) is disqualified because the negative *S2* coefficient is absolutely larger than the *SC* coefficient, which implies that increasing severity from level 1 to 2 on this dimension (self-care) decreases X_{rescaled} (as it should) yet, perversely, *increases* it from level 2 to 3. Similarly, specification (4)'s negative *UA* coefficient implies that going from level 1 to 2 (i.e., "no problems" to "moderate problems") *increases* X_{rescaled}. The inconsistencies in specifications (5) are (6) arise, broadly-speaking, from the relative magnitudes and opposite signs of the coefficients on the interaction dummies (i.e., set 3). Finally, in general, for specifications (7), (8) and (9) to produce logically consistent tariffs the coefficients on dummies *F11* to *F41* (set 4) must be monotonically decreasing, while the coefficients on *F13* to *F53* (set 5) must be increasing — which, clearly, they are not. Similar arguments apply to the six equations disqualified for the sub-sample (n=396).

The statistical significance of the constants in the preferred equations poses a challenge for their interpretation. [As noted in the previous section, theoretically they ought to be zero because when the dummies are set to zero (corresponding to state 11111, valued at unity) then $1 - X_{\text{rescaled}} = constant$, which rearranges to $X_{\text{rescaled}} = 1 - constant$.] Dolan (1997), who, like us, obtained a positive constant, suggests two possible approaches.

The first is practical: divide the equations, that is, their coefficients, by 1 – *constant* when calculating the tariffs. We found, however, that compared to the tariffs (detailed in the following section) that were generated without this transformation, the differences between the tariff values and respondents' mean values increased significantly — which we considered to be counter-productive to our objective of mirroring respondents' health state preferences. "Alternatively [and the approach we have adopted here], given that by definition the value of 11111 [is] 1, we could interpret the intercept as representing any move away from full health. Thus, [it] could represent a discontinuity in the model between level 1 and level 2 in much the same way as the '*N3*' term represents a discontinuity between level 2 and level 3." (p. 1104). Accepting this interpretation, in the following section two tariffs of health state preferences are calculated from n=919's equation (1) and n=396's equation (2).

---

[18] Applying *UA*'s negative point estimate exacerbates the tariff inconsistency problem.

6. THE TARIFFS

Two tariffs were calculated from the equations reproduced below from Tables A1 and A2 in the Appendix — here with the dependent variable reinstated as X_{rescaled}. Both tariffs for all 244 states, including unconscious (the average of the reported rescaled values), are reported in Table A3 in the Appendix. Table 7 illustrates the arithmetic by which these tariff values are calculated.

Equation (1) from n=919 (the full sample):
$X_{\text{rescaled}} = 1 - 0.2254 - 0.1026MO - 0.0894SC - 0.0346UA - 0.1046PD - 0.1127AD$

Equation (2) from n=396 (0-1 inconsistencies):
$X_{\text{rescaled}} = 1 - 0.2041 - 0.0753MO - 0.0714SC - 0.0136UA - 0.0798PD - 0.0920AD - 0.2165N3$

The relative importance to respondents (on average) of the five EQ-5D dimensions is revealed by the relative magnitudes of the dummy coefficients. In both equations anxiety/depression (*AD*) comes ahead of pain/discomfort (*PD*), followed by mobility (*MO*), then self-care (*SC*) and, finally, usual activities (*UA*). This ordering is not too dissimilar to Dolan's (1997), although his preferred equation, specification (4), recognises differences in the effects of moving between the three levels. For the movement from level 1 to 2 Dolan's ranking was: pain/discomfort, self-care, anxiety/depression, mobility, and usual activities; and for the move from level 2 to 3 it was: pain/discomfort, mobility, anxiety/depression, self-care, and usual activities.

Table 7: Illustration of how the tariff value for state 21232 is calculated using equation (2) from n=396.

	Full health (11111) =	1.000
minus	Constant term =	0.2041
minus	Mobility (*MO*): level 2 =	1 × 0.0753
minus	Self-care (*SC*): level 1 =	0 × 0.0714
minus	Usual activities (*UA*): level 2 =	1 × 0.0136
minus	Pain/discomfort (*PD*): level 3 =	2 × 0.0798
minus	Anxiety/depression (*AD*): level 2 =	1 × 0.0920
minus	Any dimension at level 3 (*N3*) =	1 × 0.2165
equals	Tariff value for state 21232 =	0.239

The 22 tariff values for which we have direct valuations from the survey are of particular interest here. A comparison (see Table 8) with the means of respondents' actual valuations provides a direct check of how well the tariffs approximate respondents' values. As summarised by the mean absolute differences, the full sample's equation (1) differs from the mean values to a greater extent than the sub-sample's equation (2). Although over two-thirds of the differences for the n=396 sample are smaller than 0.05, as a proportion of the mean value, some of them are large (e.g., –142% for state 33323). As evidenced by the reported Pearson correlation coefficients (where r = 1 for perfect positive correlation), the sub-sample's equation (2) tariff values are more closely (linearly) correlated with the corresponding mean values than the full sample's equation (1).

Although both tariffs in their entirety (Table A3) are closely related (r = 0.94), they are different.[19] In particular, the n=919 tariff values are higher than the n=396 values for 206 of the 244 states. This is explained by all of these states having at least one dimension at level 3, which for n=396's equation (2) generates an additional decrement from state 11111=1 due to the variable *N3*. On the other hand, in the the n=919 tariff unconscious has a small negative value (i.e., worse than dead) and a positive value in the other tariff (however these values are not derived from equations (1) and (2), instead they are sample averages). Finally, in both tariffs the same six states have negative values: 23333, 32333, 33133, 33233, 33323, 33333.

Table 8: Comparisons of tariff values with mean reported values.

	Full sample (n=919)			0-1 inconsistencies (n=396)		
Health state	Equation (1) tariff value	Mean value	Difference (% of mean value)	Equation (2) tariff value	Mean value	Difference (% of mean value)
11211	0.740	0.834	-0.094 (-11%)	0.782	0.796	-0.014 (-2%)
11121	0.670	0.754	-0.084 (-11%)	0.716	0.735	-0.019 (-3%)
21111	0.672	0.716	-0.044 (-6%)	0.721	0.743	-0.022 (-3%)
11112	0.662	0.769	-0.107 (-14%)	0.704	0.730	-0.026 (-4%)
12111	0.685	0.655	0.030 (5%)	0.725	0.696	0.029 (4%)
11122	0.557	0.527	0.030 (6%)	0.624	0.511	0.113 (22%)
11312	0.593	0.354	0.239 (67%)	0.460	0.369	0.091 (25%)
22222	0.331	0.35	-0.019 (-6%)	0.464	0.369	0.095 (26%)
11131	0.565	0.541	0.024 (5%)	0.420	0.496	-0.076 (-15%)
13311	0.527	0.438	0.089 (20%)	0.409	0.440	-0.031 (-7%)
11113	0.549	0.412	0.137 (33%)	0.395	0.416	-0.021 (-5%)
32211	0.445	0.221	0.224 (102%)	0.344	0.295	0.049 (17%)
11133	0.340	0.129	0.211 (164%)	0.236	0.160	0.076 (47%)
21232	0.315	0.31	0.005 (2%)	0.239	0.266	-0.027 (-10%)
22323	0.183	0.218	-0.035 (-16%)	0.142	0.146	-0.004 (-3%)
33321	0.217	0.059	0.158 (268%)	0.179	0.118	0.061 (52%)
32313	0.185	0.115	0.070 (61%)	0.146	0.146	0.000 (0%)
22233	0.113	0.152	-0.039 (-25%)	0.076	0.143	-0.067 (-47%)
32223	0.115	0.121	-0.006 (-5%)	0.080	0.124	-0.044 (-36%)
23232	0.137	0.078	0.059 (75%)	0.096	0.115	-0.019 (-16%)
33323	-0.009	0.032	-0.041 (-127%)	-0.005	0.012	-0.017 (-142%)
33333	-0.113	-0.05	-0.066 (141%)	-0.085	-0.074	-0.011 (15%)
mean absolute difference:			0.082 (53%)			0.041 (23%)
Pearson's correlation coefficient:	r = 0.92			r = 0.98		

---

[19] However, it ought to be acknowledged that the confidence intervals (not reported) around each estimate probably blur most of the differences, as well as the rankings of states.

7. DISCUSSION AND CONCLUDING REMARKS

As explained in the Data section, the health state valuations in the present study were elicited via a visual analogue scale (VAS). Several caveats apply to the use and interpretation of values derived in this manner. One arises from the requirement that the duration of the state being valued be specified. Respondents were asked to “imagine that it will last for one year. What happens after that is not known and should not be taken in to account.” Gudex and Dolan (1995) have shown that respondents’ valuations can be affected by the period of duration that is specified.

Another concern is the possibility that respondents may be unwilling to use values at or near the extremes of the VAS continuum (0 and 100) for scoring states. Badia et al. (1999b), comparing valuation approaches for EQ-5D health states, found that considerably more states were rated worse than dead using time trade-off (TTO) than using VAS and that “... VAS values were compressed into a much tighter valuation space than TTO values” (p. 309). On the other hand, they confirmed the general finding in the literature that, notwithstanding the problems associated with VAS, it is more feasible and reliable than TTO.

Compared to both the VAS and TTO, the standard gamble approach is regarded by some to be the ‘gold standard’, because of its theoretical foundations and indirect approach to valuations (notwithstanding its implementation shortcomings). It is well known that VAS values are theoretically inconsistent with the ‘utility under uncertainty’ provisions of von Neumann and Morgenstern expected utility theory (Drummond et al. 1997, Sloan 1995). In practical terms, the values generated for given health states are not invariant to these three methods (Sloan 1995). Hence, comparisons between our tariff and the UK tariff published by Dolan et al. (1995) are of little relevance, since the latter is based on TTO interviews.[20]

At just 50%, the response rate to the questionnaire on which the present study is based may be considered low, although it is close to that experienced in similar research internationally (see Brooks 1996, Bjork and Norinder 1999). This is likely to be related to the responder burden imposed by the valuation task. Although self-completed VAS-based questionnaires are widely used internationally, and a relatively inexpensively means of undertaking large-scale valuation exercises, it is possible that, as Bjork and Norinder (p. 123) have suggested, health state valuation is “... too complicated for this approach.”

A related concern is the quality of the data generated by the questionnaire. At the start of the analysis almost a third of responses had to be rejected for a variety of reasons (the most common

---

[20] A study is currently underway in the UK using a survey instrument and research protocol similar to that reported here, which will enable direct comparisons to be made.

being failure to score state 11111 or dead on the VAS — see Table 2). Fortunately, exclusions on these grounds did not appear to seriously compromise the representativeness of the sample in terms of broad demographic characteristics (see Table 3). However, whether individuals who did not respond to the questionnaire, or who did so inadequately, differed systematically *in terms of their health state preferences* from those who did respond is unknown.

In addition, and a central theme of the paper, 80% of the usable responses exhibited at least one logical inconsistency (see Table 4). Again, given the responder burden, this and the previous statistic are perhaps not surprising. But they do give rise to questions about the ability of ordinary members of the public to undertake the valuation task unaided. The inconsistency rate for the present sample is considerably higher than the 9.3% rate reported by Dolan and Kind (1996) for a self-completed VAS postal survey in the UK and also higher than the 26% reported by Badia et al. (1999a) for an interview-administered VAS on a Catalan sample. The highest reported rate of inconsistency is 88% from a self-completed postal VAS survey in the US (Johnson et al. 1998). However, direct comparisons of these rates is problematic given the different approaches to data exclusions *prior* to analysis; for example, Badia et al. (1999a) exclude some inconsistency 'types' which are included in our analysis.[21] While Badia et al. conclude that the inconsistencies they analysed did not affect rankings in their final tariff of values, this was not the case for the present data set.

There are therefore unresolved methodological questions relating to the treatment of logical inconsistencies in respondents' valuations. Our approach was to use two samples at opposite extremes — to admit all inconsistencies (n=919) and to admit none or one (n=396). As discussed in the previous section, the resulting tariffs are different. Therefore we believe that there is merit in reporting the effect on results of methodological decisions such as this that rest heavily on researcher judgments. The air of precision leant to tariffs by the listing of a single value per state may be misleading because of the possibility of alternative tariffs.

An implication for the practice of economic evaluation is that health state values used in CUA should be subjected to careful sensitivity analysis — gauging the effects of using alternative tariffs — in the same way as any other uncertain but important variable. We recommend that New Zealand analysts proposing to use the values reported here test the effects on their results of using values from *both* tariffs and consider these alongside the results produced when the UK tariff (or others) is employed.

---

[21] Badia et al. excluded individuals "... if they valued any health state higher than the logically best health state (11111)" (p. 944). They did not provide their reasons for treating these inconsistencies differently from the ones they analysed, nor did they report how many exclusions were on these grounds.

We note that the existing literature exploring issues arising from inconsistency in health state valuation (e.g. Dolan and Kind 1996, Ohinmaa and Sintonen 1998, Badia et al. 1999a) appears to focus exclusively on data generated from the EQ-5D health state classification system. Yet other systems are also likely to suffer from this problem, especially those with more dimensions and/or levels, such as the HUI mentioned at the beginning of the paper.

A significant research territory remains to be explored in New Zealand in relation to health state valuations. Data generated for this study will allow us, in future papers, to explore the extent to which the EQ-5D health state classification system captures all the aspects of health considered important by our sample of New Zealanders, and to investigate the extent to which valuations can be explained by their socio-demographic characteristics. A key issue to explore in future research is the health state valuations of Maori, and the adequacy of the EQ-5D in describing Maori health perspectives. Furthermore, the features of the data set discussed above suggest that a study of the New Zealand general public using interview-based, indirect approaches (as an alternative to the methods used in the current study) would be worthwhile. Given the importance of the resource allocation decisions this sort of research may inform, there is an urgent requirement for the development of a base of New Zealand evidence on these and wider questions about health state preferences.

Appendix

Table A1: Estimates of specifications (1) - (9) for the full sample (n=919).

Dummy variable					Specification				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MO	0.1026	0.0888	0.0705	0.0882	0.2804	0.1751		0.0496♣	0.0915
	(8.71)	(7.57)	(3.60)	(4.53)	(10.36)	(7.58)		(0.92)	(7.04)
SC	0.0894	0.0881	0.1768	0.1571	0.3506	0.2138		0.0749♣	0.1216
	(6.30)	(6.26)	(8.14)	(7.26)	(10.98)	(7.95)		(1.44)	(7.01)
UA	0.0346	-0.0187♣	0.0211♣	-0.0522	0.1570	0.0219♣		-0.0887	-0.0095♣
	(2.60)	(-1.36)	(1.01)	(-2.42)	(4.96)	(0.82)		(-2.05)	(-0.64)
PD	0.1046	0.0621	0.0763	0.0910	0.2560	0.1514		0.0293♣	0.0816
	(10.67)	(6.06)	(4.42)	(5.30)	(8.42)	(8.90)		(0.61)	(6.83)
AD	0.1127	0.0788	0.0959	0.1171	0.2480	0.1654		0.0516♣	0.1066
	(11.14)	(7.60)	(5.20)	(6.36)	(9.25)	(9.24)		(1.14)	(8.33)
M2			0.0505♣	0.0015♣	0.3881				
			(1.38)	(0.04)	(3.65)				
S2			-0.2099	-0.1671	0.1082♣				
			(-5.27)	(-4.22)	(1.28)				
U2			0.0623♣	0.0870	0.5758				
			(1.63)	(2.30)	(6.44)				
P2			0.1133	-0.0240♣	-0.0748♣				
			(3.73)	(-0.75)	(-1.23)				
A2			0.0009♣	-0.1126	0.0949♣				
			(0.03)	(-2.94)	(1.90)				
N3		0.2579		0.2800					
		(13.12)		(12.74)					
MOSC					-0.3658	-0.0480♣			
					(-2.80)	(-0.59)			
MOUA					0.1718♣	0.0791♣			
					(1.77)	(1.06)			
MOPD					0.1548♣	0.2288♣			
					(0.98)	(1.80)			
MOAD					-0.2767	-0.1801			
					(-2.68)	(-2.37)			
SCUA					-0.2863	-0.0060♣			
					(-6.19)	(-0.33)			
SCPD					-0.0211♣	-0.0079♣			
					(-0.79)	(-0.36)			
SCAD					0.3924	-0.0344♣			
					(3.55)	(-0.44)			
UAPD					-0.1442	-0.3034			
					(-0.92)	(-2.41)			
UAAD					-0.2416	0.1071			
					(-4.02)	(3.36)			
PDAD					-0.0414	0.0319			
					(-2.46)	(2.55)			
F11							-0.1382	-0.1253♣	
							(-2.41)	(-1.72)	
F21							-0.0613♣	-0.0597	
							(-1.38)	(-0.61)	
F31							-0.1422	-0.0291♣	
							(-5.88)	(-0.20)	
F41							-0.3606	-0.2949♣	
							(-14.46)	(-1.61)	
F13							0.2348	0.2423	0.2455
							(9.05)	(5.01)	(11.05)
F23							0.2784	0.1738♣	0.1782
							(10.90)	(1.80)	(5.50)
F33							0.4437	0.4250	0.2397
							(7.19)	(2.87)	(5.40)
F43							0.3470	0.2422♣	0.0775♣
							(9.68)	(1.29)	(1.49)
F53							0.4321	0.2959♣	0.0799♣
							(15.14)	(1.27)	(1.54)
constant	0.2254	0.2102	0.2263	0.1779	-0.0032♣	0.1039	0.6113	0.5142	0.1799
	(8.54)	(7.96)	(8.26)	(6.44)	(-0.09)	(3.41)	(18.32)	(2.20)	(6.51)
R^2	0.08	0.09	0.08	0.09	0.09	0.09	0.09	0.09	0.09

Notes: Figures in parentheses are t-values and ♣ denotes ***in****significance at the 5% level.*

Table A2: Estimates of specifications (1) - (9) for the sample with 0-1 inconsistencies (n=396).

Dummy variable					Specification				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MO	0.0895	0.0753	0.0536	0.0662	0.2256	0.1345		-0.0144$^{\clubsuit}$	0.0740
	(21.11)	(19.66)	(7.81)	(10.68)	(27.56)	(17.90)		(-0.87)	(17.46)
SC	0.0719	0.0714	0.1340	0.1165	0.2751	0.1544		0.0050$^{\clubsuit}$	0.0953
	(14.07)	(15.55)	(17.55)	(16.87)	(28.46)	(17.59)		(0.31)	(16.87)
UA	0.0587	0.0136	0.0416	-0.0183	0.1733	0.0543		-0.0803	0.0172
	(12.15)	(2.99)	(5.67)	(-2.67)	(18.11)	(6.26)		(-5.98)	(3.59)
PD	0.1163	0.0798	0.0992	0.1091	0.2434	0.1581		0.0051$^{\clubsuit}$	0.0921
	(33.03)	(23.83)	(16.38)	(19.93)	(26.59)	(28.36)		(0.35)	(23.94)
AD	0.1180	0.0920	0.1031	0.1219	0.2530	0.1615		0.0229$^{\clubsuit}$	0.1106
	(32.43)	(27.36)	(15.89)	(20.73)	(31.29)	(27.12)		(1.63)	(26.63)
M2			0.0680	0.0258	0.3804				
			(5.34)	(2.23)	(12.02)				
S2			-0.1547	-0.1135	0.0859				
			(-11.00)	(-8.90)	(3.40)				
U2			0.0595	0.0757	0.5257				
			(4.41)	(6.21)	(19.37)				
P2			0.0809	-0.0331	-0.0256$^{\clubsuit}$				
			(7.56)	(-3.23)	(-1.40)				
A2			-0.0006$^{\clubsuit}$	-0.0909	0.0462				
			(-0.048)	(-7.45)	(3.04)				
N3		0.2165		0.2338					
		(33.64)		(33.02)					
MOSC					-0.2177	0.0211$^{\clubsuit}$			
					(-5.52)	(0.80)			
MOUA					0.0399$^{\clubsuit}$	0.0087$^{\clubsuit}$			
					(1.36)	(0.36)			
MOPD					-0.0782$^{\clubsuit}$	0.0713$^{\clubsuit}$			
					(-1.65)	(1.75)			
MOAD					-0.1264	-0.0834			
					(-4.05)	(-3.37)			
SCUA					-0.2454	0.0004$^{\clubsuit}$			
					(-17.58)	(0.06)			
SCPD					-0.0412	-0.0225			
					(-5.10)	(-3.16)			
SCAD					0.2976	-0.0539			
					(8.92)	(-2.11)			
UAPD					0.1039	-0.1201			
					(2.23)	(-2.97)			
UAAD					-0.2587	0.0607			
					(-14.28)	(5.75)			
PDAD					-0.0514	0.0192			
					(-9.85)	(4.64)			
F11							-0.1090	-0.1364	
							(-6.06)	(-6.16)	
F21							-0.1396	-0.1240	
							(-9.68)	(-4.08)	
F31							-0.1478	-0.1833	
							(-18.84)	(-4.05)	
F41							-0.3444	-0.4142	
							(-42.41)	(-7.29)	
F13							0.2433	0.2544	0.2160
							(28.39)	(16.87)	(29.80)
F23							0.2771	0.2558	0.1583
							(33.50)	(8.59)	(15.03)
F33							0.3734	0.4786	0.1998
							(19.38)	(10.47)	(14.05)
F43							0.3793	0.4330	0.1141
							(32.36)	(7.46)	(6.74)
F53							0.4679	0.5155	0.1094
							(50.60)	(7.16)	(6.50)
constant	0.2156	0.2041	0.2219	0.1832	0.0278	0.1258	0.6055	0.6814	0.1856
	(28.73)	(27.68)	(27.99)	(23.44)	(2.79)	(14.12)	(60.90)	(9.42)	(23.69)
R^2	0.68	0.71	0.68	0.72	0.73	0.71	0.71	0.73	0.72

*Notes: Figures in parentheses are t-values and ♣ denotes **in**significance at the 5% level*

TABLE A3: TWO TARIFFS OF HEALTH STATE PREFERENCES OF NEW ZEALANDERS

state	n=919 Eq. (1)	n=396 Eq. (2)	*difference*	state	n=919 Eq. (1)	n=396 Eq. (2)	*difference*	state	n=919 Eq. (1)	n=396 Eq. (2)	*difference*
11111	1.000	1.000	*0.000*	12133	0.251	0.164	*0.086*	13232	0.239	0.171	*0.068*
11112	0.662	0.704	*-0.042*	12211	0.651	0.711	*-0.060*	13233	0.127	0.079	*0.047*
11113	0.549	0.395	*0.154*	12212	0.538	0.619	*-0.081*	13311	0.527	0.409	*0.117*
11121	0.670	0.716	*-0.046*	12213	0.425	0.310	*0.115*	13312	0.414	0.317	*0.096*
11122	0.557	0.624	*-0.067*	12221	0.546	0.631	*-0.085*	13313	0.301	0.225	*0.076*
11123	0.445	0.316	*0.129*	12222	0.433	0.539	*-0.106*	13321	0.422	0.330	*0.092*
11131	0.565	0.420	*0.146*	12223	0.321	0.231	*0.090*	13322	0.309	0.238	*0.072*
11132	0.453	0.328	*0.125*	12231	0.441	0.335	*0.106*	13323	0.197	0.146	*0.051*
11133	0.340	0.236	*0.104*	12232	0.329	0.243	*0.086*	13331	0.317	0.250	*0.067*
11211	0.740	0.782	*-0.042*	12233	0.216	0.151	*0.065*	13332	0.205	0.158	*0.047*
11212	0.627	0.690	*-0.063*	12311	0.616	0.481	*0.135*	13333	0.092	0.066	*0.026*
11213	0.515	0.382	*0.133*	12312	0.503	0.389	*0.114*	21111	0.672	0.721	*-0.049*
11221	0.635	0.703	*-0.067*	12313	0.391	0.297	*0.094*	21112	0.559	0.629	*-0.069*
11222	0.523	0.611	*-0.088*	12321	0.511	0.401	*0.110*	21113	0.447	0.320	*0.127*
11223	0.410	0.302	*0.108*	12322	0.399	0.309	*0.090*	21121	0.567	0.641	*-0.073*
11231	0.531	0.406	*0.124*	12323	0.286	0.217	*0.069*	21122	0.455	0.549	*-0.094*
11232	0.418	0.314	*0.104*	12331	0.407	0.321	*0.085*	21123	0.342	0.240	*0.102*
11233	0.305	0.222	*0.083*	12332	0.294	0.229	*0.065*	21131	0.463	0.345	*0.118*
11311	0.705	0.552	*0.153*	12333	0.181	0.137	*0.044*	21132	0.350	0.253	*0.098*
11312	0.593	0.460	*0.132*	13111	0.596	0.437	*0.159*	21133	0.237	0.160	*0.077*
11313	0.480	0.368	*0.112*	13112	0.483	0.345	*0.139*	21211	0.637	0.707	*-0.070*
11321	0.601	0.472	*0.128*	13113	0.371	0.253	*0.118*	21212	0.525	0.615	*-0.090*
11322	0.488	0.380	*0.108*	13121	0.491	0.357	*0.134*	21213	0.412	0.306	*0.106*
11323	0.375	0.288	*0.087*	13122	0.379	0.265	*0.114*	21221	0.533	0.627	*-0.094*
11331	0.496	0.393	*0.103*	13123	0.266	0.173	*0.093*	21222	0.420	0.535	*-0.115*
11332	0.383	0.301	*0.083*	13131	0.387	0.277	*0.110*	21223	0.307	0.227	*0.081*
11333	0.271	0.209	*0.062*	13132	0.274	0.185	*0.089*	21231	0.428	0.331	*0.097*
12111	0.685	0.725	*-0.039*	13133	0.161	0.093	*0.068*	21232	0.315	0.239	*0.077*
12112	0.573	0.633	*-0.060*	13211	0.561	0.423	*0.138*	21233	0.203	0.147	*0.056*
12113	0.460	0.324	*0.136*	13212	0.449	0.331	*0.118*	21311	0.603	0.477	*0.126*
12121	0.581	0.645	*-0.064*	13213	0.336	0.239	*0.097*	21312	0.490	0.385	*0.105*
12122	0.468	0.553	*-0.085*	13221	0.457	0.343	*0.113*	21313	0.377	0.293	*0.085*
12123	0.355	0.244	*0.111*	13222	0.344	0.251	*0.093*	21321	0.498	0.397	*0.101*
12131	0.476	0.348	*0.128*	13223	0.231	0.159	*0.072*	21322	0.385	0.305	*0.080*
12132	0.363	0.256	*0.107*	13231	0.352	0.264	*0.088*	21323	0.273	0.213	*0.060*

state	n=919 Eq. (1)	n=396 Eq. (2)	*difference*	state	n=919 Eq. (1)	n=396 Eq. (2)	*difference*	state	n=919 Eq. (1)	n=396 Eq. (2)	*difference*
21331	0.394	0.317	*0.076*	23131	0.284	0.202	*0.082*	31231	0.326	0.256	*0.070*
21332	0.281	0.225	*0.056*	23132	0.171	0.110	*0.062*	31232	0.213	0.164	*0.049*
21333	0.168	0.133	*0.035*	23133	0.059	0.018	*0.041*	31233	0.100	0.072	*0.029*
22111	0.583	0.649	*-0.067*	23211	0.459	0.348	*0.111*	31311	0.500	0.402	*0.099*
22112	0.470	0.557	*-0.087*	23212	0.346	0.256	*0.090*	31312	0.388	0.310	*0.078*
22113	0.357	0.249	*0.109*	23213	0.233	0.164	*0.070*	31313	0.275	0.218	*0.057*
22121	0.478	0.569	*-0.091*	23221	0.354	0.268	*0.086*	31321	0.396	0.322	*0.074*
22122	0.365	0.477	*-0.112*	23222	0.241	0.176	*0.065*	31322	0.283	0.230	*0.053*
22123	0.253	0.169	*0.084*	23223	0.129	0.084	*0.045*	31323	0.170	0.138	*0.032*
22131	0.373	0.273	*0.100*	23231	0.249	0.188	*0.061*	31331	0.291	0.242	*0.049*
22132	0.261	0.181	*0.080*	23232	0.137	0.096	*0.041*	31332	0.178	0.150	*0.028*
22133	0.148	0.089	*0.059*	23233	0.024	0.004	*0.020*	31333	0.066	0.058	*0.008*
22211	0.548	0.636	*-0.088*	23311	0.424	0.334	*0.090*	32111	0.480	0.357	*0.123*
22212	0.435	0.544	*-0.108*	23312	0.311	0.242	*0.069*	32112	0.367	0.265	*0.102*
22213	0.323	0.235	*0.088*	23313	0.199	0.150	*0.049*	32113	0.255	0.173	*0.081*
22221	0.443	0.556	*-0.112*	23321	0.319	0.254	*0.065*	32121	0.375	0.278	*0.098*
22222	0.331	0.464	*-0.133*	23322	0.207	0.162	*0.044*	32122	0.263	0.186	*0.077*
22223	0.218	0.155	*0.063*	23323	0.094	0.070	*0.024*	32123	0.150	0.094	*0.057*
22231	0.339	0.260	*0.079*	23331	0.215	0.175	*0.040*	32131	0.271	0.198	*0.073*
22232	0.226	0.168	*0.059*	23332	0.102	0.083	*0.020*	32132	0.158	0.106	*0.052*
22233	0.113	0.076	*0.038*	23333	-0.011	-0.009	*-0.001*	32133	0.045	0.014	*0.032*
22311	0.513	0.406	*0.108*	31111	0.569	0.429	*0.141*	32211	0.445	0.344	*0.102*
22312	0.401	0.313	*0.087*	31112	0.457	0.337	*0.120*	32212	0.333	0.252	*0.081*
22313	0.288	0.221	*0.067*	31113	0.344	0.245	*0.099*	32213	0.220	0.160	*0.060*
22321	0.409	0.326	*0.083*	31121	0.465	0.349	*0.116*	32221	0.341	0.264	*0.077*
22322	0.296	0.234	*0.062*	31122	0.352	0.257	*0.095*	32222	0.228	0.172	*0.056*
22323	0.183	0.142	*0.042*	31123	0.239	0.165	*0.075*	32223	0.115	0.080	*0.035*
22331	0.304	0.246	*0.058*	31131	0.360	0.269	*0.091*	32231	0.236	0.184	*0.052*
22332	0.191	0.154	*0.038*	31132	0.248	0.177	*0.070*	32232	0.124	0.092	*0.031*
22333	0.079	0.062	*0.017*	31133	0.135	0.085	*0.050*	32233	0.011	0.000	*0.011*
23111	0.493	0.361	*0.132*	31211	0.535	0.415	*0.120*	32311	0.411	0.330	*0.081*
23112	0.381	0.269	*0.111*	31212	0.422	0.323	*0.099*	32312	0.298	0.238	*0.060*
23113	0.268	0.177	*0.091*	31213	0.309	0.231	*0.078*	32313	0.185	0.146	*0.039*
23121	0.389	0.282	*0.107*	31221	0.430	0.335	*0.095*	32321	0.306	0.250	*0.056*
23122	0.276	0.190	*0.086*	31222	0.318	0.243	*0.074*	32322	0.194	0.158	*0.035*
23123	0.163	0.097	*0.066*	31223	0.205	0.151	*0.053*	32323	0.081	0.066	*0.014*

state	n=919 Eq. (1)	n=396 Eq. (2)	*difference*
32331	0.202	0.171	*0.031*
32332	0.089	0.079	*0.010*
32333	-0.024	-0.013	*-0.010*
33111	0.391	0.286	*0.105*
33112	0.278	0.194	*0.084*
33113	0.165	0.102	*0.063*
33121	0.286	0.206	*0.080*
33122	0.173	0.114	*0.059*
33123	0.061	0.022	*0.039*
33131	0.181	0.126	*0.055*
33132	0.069	0.034	*0.034*
33133	-0.044	-0.058	*0.014*
33211	0.356	0.272	*0.084*
33212	0.243	0.180	*0.063*
33213	0.131	0.088	*0.042*
33221	0.251	0.193	*0.059*
33222	0.139	0.101	*0.038*
33223	0.026	0.009	*0.018*
33231	0.147	0.113	*0.034*
33232	0.034	0.021	*0.013*
33233	-0.079	-0.071	*-0.007*
33311	0.321	0.259	*0.063*
33312	0.209	0.167	*0.042*
33313	0.096	0.075	*0.021*
33321	0.217	0.179	*0.038*
33322	0.104	0.087	*0.017*
33323	-0.009	-0.005	*-0.004*
33331	0.112	0.099	*0.013*
33332	0.000	0.007	*-0.008*
33333	-0.113	-0.085	*-0.028*
unconscious[*]	-0.004	0.032	*-0.036*

---

[*] Since unconscious is incapable of being represented by dummy variables its tariff value is the mean value from the sample.

REFERENCES

Ashton, T., Cumming, J., Devlin, N. (1999), *Prioritising Health and Disability Support Services: Principles, Processes and Problems. A Review of the HFA Prioritisation Process*. Wellington: National Health Committee.

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

Badia, X., Monserrat, S., Roset, M., Herdman, M. (1999b), Feasibility, validity and test-retest reliability of scaling methods for health states: the visual analogue scale and the time trade-off, *Quality of Life Research*, 8, 303-310.

Bjork, S., Norinder, A. (1999), The weighting exercise for the Swedish version of the EuroQol, *Health Economics,* 8, 117-126.

Brooks, R., EuroQol Group (1996), EuroQol: the current state of play, *Health Policy*, 37, 53-72.

Devlin, N., Williams, A.H. (1999), Valuing quality of life: results for New Zealand health professionals, *New Zealand Medical Journal*, 112, 68-71.

Devlin, N., Ashton, T., Cumming, J. (1999), Rationing health care: how should the HFA proceed? *New Zealand Medical Journal* , 112, 369-370.

Dolan, P. (1999) Whose preferences count? *Medical Decision Making,* 19, 482-486.

Dolan, P., Kind, P. (1996), Inconsistency and health state valuations, *Social Science and Medicine*, 42, 4, 609-615.

Dolan, P. (1997), Modeling valuations for EuroQol health states, *Medical Care*, 35, 11, 1095-1108.

Dolan, P., Gudex, C., Kind, P., Williams, A. (1994), *The Measurement and Valuation of Health. First Report on the Main Survey*. The MVH Group. Centre for Health Economics, University of York.

Dolan, P., Gudex, C., Kind, P., Williams, A. (1995), *A Social Tariff for EuroQol: Results from a UK General Population Survey,* Discussion Paper 138, Centre for Health Economics, University of York.

Dolan, P., Kind, P. (1996), Inconsistency and health state valuations, *Social Science and Medicine*, 42, 4, 609-615.

Drummond, M.F., O'Brien, B., Stoddart, G.L., Torrance, G.W. (1997), *Methods for the Economics Evaluation of Health Care Programmes*, (2nd edition), Oxford Medical Publications.

Durie, M. (1998), *Whaiora. Maori Health Development*, Oxford University Press, 2nd edition.

Greene, W.H. (1991), *LIMDEP User's Manual and Reference Guide, Version 6*, Econometric Software, Inc. New York.

Gudex, C., Dolan, P. (1995), *Valuing health states: the effect of duration*, Discussion Paper 143, Centre for Health Economics, University of York.

Health Funding Authority (1998), *How Shall We Prioritise Our Health and Disability Support Services? A Discussion Paper*. Wellington: HFA.

Johnson, J.A., Ergo, A., Coons, S.J. (1998), Valuation of the EuroQol (EQ-5D) in an adult US sample, *Pharmacoeconomics* 13, 421-433.

Ohinmaa, A., Sintonen, H. (1998), Inconsistencies and modeling of the Finnish EuroQol (EQ-5D) preference values, paper presented to the 15th EuroQol Plenary Meeting, Hanover, 1st and 2nd October.

PHARMAC (1998), *A Prescription for Pharmacoeconomic Analysis*, Wellington: PHARMAC.

Sloan, F.A. (editor) (1995), *Valuing Health Care*, Cambridge University Press.

Statistics New Zealand (1997), *1996 Census of Population and Dwellings. National Summary*, Wellington: Statistics New Zealand.

Weinstein, M.C., Siegel, J.E., Gold, M.R., Kamlet, M.S., Russell, L.B; for the Panel on Cost-Effectiveness in Health and Medicine (1996), Recommendation of the Panel on Cost-Effectiveness in Health and Medicine, *Journal of the American Medical Association*, 276, 1253-8.