EQ-5D CONCEPTS AND METHODS
EQ-5D concepts and methods: a developmental history

Edited by

PAUL KIND
University of York, U.K.

RICHARD BROOKS
University of Strathclyde, Strathclyde, Scotland

and

ROSALIND RABIN
EuroQol Group Business Management, The Netherlands
A collection of papers representing the collective intellectual enterprise of the EuroQol Group

Presented to
Alan Williams

by his friends
in the EuroQol Group in Rotterdam,
October 1997

on the happy coincidence of his 70th birthday
and the 10th year of the Group’s existence
Table of contents

List of contributors ix

Foreword xi

List of tables and appendices xv

List of figures xxv

1. The EuroQol Instrument 1
   Alan Williams

2. The descriptive system of the EuroQol Instrument 19
   Claire Gudex

3. The number of levels in the descriptive system 29
   Heleen van Agt and Gouke Bonsel

4. First steps to assessing semantic equivalence of the EuroQol Instrument: Results of a questionnaire survey to members of the EuroQol Group 35
   Julia Fox-Rushby

5. Comparing general health related quality of life (HRQoL) questionnaires; EuroQol, Sickness Impact Profile and Rosser Index 53
   Stefan Björk and Ulf Persson

6. Influence of self-rated health and related variables on EuroQol valuation of health states in a Spanish population 63
   Xavier Badia, Esteve Fernandez and Andreu Segura

7. Observations on one hundred students filling in the EuroQol questionnaire 81
   Jan Busschbach, Dick Hessing and Frank de Charro

8. Eliciting EuroQol descriptive data and utility scale values from inpatients 91
   Caroline Selai and Rachel Rosser
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Test-retest reliability of health state valuations collected with the EuroQol questionnaire</td>
<td>Heleen van Agt, Marie-Louise Essink-Bot, Paul Krabbe and Gouke Bonsel</td>
<td>109</td>
</tr>
<tr>
<td>10</td>
<td>Hypothetical valuations of health states versus patients’ self-ratings</td>
<td>Erik Nord, Xavier Badia, Montserrat Rue and Harri Sintonen</td>
<td>125</td>
</tr>
<tr>
<td>11</td>
<td>Inconsistency and health state valuations</td>
<td>Paul Dolan and Paul Kind</td>
<td>139</td>
</tr>
<tr>
<td>12</td>
<td>Issues in the harmonisation of valuation and modeling</td>
<td>Paul Krabbe, Frank de Charro and Marie-Louise Essink-Bot</td>
<td>147</td>
</tr>
<tr>
<td>13</td>
<td>Estimating a parametric relation between health description and health valuation using the EuroQol Instrument</td>
<td>Ben van Hout and Joseph McDonnell</td>
<td>157</td>
</tr>
<tr>
<td>14</td>
<td>Some considerations concerning negative values for EQ-5D health states</td>
<td>Frank de Charro, Jan Busschbach, Marie-Louise Essink-Bot, Ben van Hout and Paul Krabbe</td>
<td>171</td>
</tr>
<tr>
<td>15</td>
<td>Health states considered worse than 'being dead'</td>
<td>Stefan Björk and Rikard Althin</td>
<td>181</td>
</tr>
<tr>
<td>16</td>
<td>The effect of duration on the values given to the EuroQol states</td>
<td>Arto Ohinmaa and Harri Sintonen</td>
<td>191</td>
</tr>
<tr>
<td>17</td>
<td>Applying paired comparisons models to EQ-5D valuations - deriving TTO utilities from ordinal preference data</td>
<td>Paul Kind</td>
<td>201</td>
</tr>
<tr>
<td>18</td>
<td>The use and usefulness of the EuroQol EQ-5D: preliminary results from an international survey</td>
<td>Rosalind Rabin, Paul Kind and Frank de Charro</td>
<td>221</td>
</tr>
<tr>
<td>19</td>
<td>Not a quick fix</td>
<td>Martin Buxton</td>
<td>235</td>
</tr>
<tr>
<td></td>
<td>Postscript</td>
<td>Alan Williams</td>
<td>239</td>
</tr>
</tbody>
</table>
List of contributors

Heleen van Agt, Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands

Rikard Althin, Pfizer UK, Sandwich, Kent, United Kingdom

Xavier Badia, Health Outcomes Research Europe, Barcelona, Spain

Stefan Björk, Global Health Economics, Novo Nordisk A/S, Bagsvaerd, Denmark

Gouke Bonsel, Academic Medical Centre, Amsterdam, The Netherlands

Jan Busschbach, Institute for Medical Psychology & Psychotherapy, Erasmus Medical Center, Rotterdam, The Netherlands

Martin Buxton, Health Economics Research Group, Brunel University, London, United Kingdom

Frank de Charro, Centre for Health Policy and Law, Erasmus University Rotterdam, Rotterdam, The Netherlands

Marie-Louise Essink-Bot, Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands

Estève Fernandez, Institut Català d'Oncologia (ICO), Barcelona, Spain

Julia Fox-Rushby, Health Economics Research Group, Brunel University, London, United Kingdom

Claire Gudex, Centre for Applied Health Services Research and Technology Assessment, University of Southern Denmark, Odense, Denmark

†Dick Hessing, Faculty of Law, Erasmus University Rotterdam, Rotterdam, The Netherlands

Ben van Hout, University Medical Centre, Utrecht, The Netherlands

Paul Kind, Outcomes Research Group, Centre for Health Economics, University of York, York, United Kingdom

Paul Krabbe, Department of Medical Technology Assessment, University Medical Centre Nijmegen, Nijmegen, The Netherlands

Joseph McDonnell, Institute for Medical Technology Assessment, Erasmus Medical Center, Rotterdam, The Netherlands
Erik Nord, Norwegian Institute of Public Health, Oslo, Norway

Ulf Persson, The Swedish Institute for Health Economics, Lund, Sweden

Arto Ohinmaa, Department of Public Health, University of Alberta, Edmonton, Canada

Rosalind Rabin, EuroQol Group Business Management, Rotterdam, The Netherlands

Montserrat Rue, Catalan Health Department, Lleida, Spain

†Rachel Rosser, Department of Psychiatry, University College London Medical School, London, United Kingdom

Andreu Segura, Institut d'Estudis de la Salut, Barcelona, Spain

Caroline Selai, Institute of Neurology, University College London, London, United Kingdom

Harri Sintonen, Department of Public Health, University of Helsinki, Helsinki, Finland

Alan Williams, Centre for Health Economics, University of York, York, United Kingdom
Science today makes progress through the imaginative harvesting of knowledge generated by the many, rather than as the result of the isolated endeavours of the lone researcher. Innovations in the physical sciences from the development of nuclear technologies to the laser, have involved research teams working collectively. Collaboration is the rule rather than the exception. In the social sciences this model is all but reversed. Here it is not uncommon to encounter the solitary enthusiast, relishing an independence of spirit and pursuing their own private research agenda. All the more surprising then that a group of researchers from several different disciplines, should have come together in the late 1980s with nothing more substantial on the agenda than that they share their thoughts on the topic of measuring the value of health, or more specifically, on the way that the value of health might vary across different countries. Few scientific enterprises can have begun as cautiously or uncertainly. Few can have developed a cohesion and dynamism that lasted decades and continues to drive ahead after long years of scientific endeavour. Such is the good fortune that befell those of us who came together to form what was later to be known as the Euro- Qol Group. The Group's creation is principally due to the shared professional association of its members with one man, an economist by training and a visionary academic by inclination and temperament - Alan Williams. It was his catalytic influence that encouraged us all to participate in an initial exploratory session in Rotterdam in 1987. No one present had the remotest idea about the remarkable journey of discovery that lay ahead, or that a simple but effective health status measurement technology was to be the pay-off for the industry we undertook together.

The Group's founder members came from disparate disciplines and different environments - from health economics, medicine, sociology and psychology, from academia, health care, public health and government. Some, such as Rachel Rosser and Harri Sintonen, had already published their own independent measures of health status. From the outset they adopted a working practice based on openness and collegiality. All participants were credited with equal status. The way ahead lay through joint discussion of ideas and the formulation of methods for dealing with concepts and problems that obstructed progress. So as to explore the variation in values for health the Group needed a standard mechanism for describing health. Initially defined in terms of 6 dimensions, this classification was streamlined in 1993, taking on the form that is known today as EQ-5D. Since the Group was interested in exploring variation in values for health it logically followed that a standard valuation method was also required. Early on the Group adopted visual analogue scaling (VAS) utilizing a vertical 20cm scale calibrated, and sometimes referred to, as a "thermometer". A standard format was adopted too in regards to the design of a valuation questionnaire for use in postal surveys of the general population. The questionnaire elicited values for 16 health states presented in two groups of eight on separate pages. Health states were described as composites based on the standard descriptive system. Values for health
states were indicated by drawing a line from the text box containing the health state description to the VAS "thermometer". Since both the concept of valuing health and the chosen technique for recording those values were likely to be foreign to most individuals presented with the task, it seemed obvious that some limited practice would be required. Two preliminary pages in the valuation questionnaire were designed to familiarize respondents with the descriptive system and the "thermometer" scale. It was only after early studies, designed to test practical issues of feasibility that it emerged that these two pages were capturing specific (and intriguing) information on respondent self-assessed health status. After much further experimentation and some limited refinement, the current version of EQ-5D was published, following a 1993 moratorium on modification that has largely held until the present.

The reader is invited to sample this collection of research memorabilia with a degree of caution. Despite the dominant rationale of the EuroQol Group in developing the technology that now bears the label EQ-5D, the issues and themes to which it has given its attention speaks to a wider, more generalized audience. For the most part these issues are not in fact particular to EQ-5D, rather they are fundamental in nature and are expressed here as a local outcropping of a more substantial substrata. It must be recalled that, with the occasional exception, these papers were addressed to other members of the EuroQol Group as a contribution to a dynamic research agenda. There is therefore, an assumed level of knowledge that marks out these papers for a specialist audience. However, there is a high degree of transparency in their content so that the interested reader may easily transfer material into their own preferred area of competence. For those with a forensic inclination, the material contained within this volume might appear to offer some scope for exploitation. The EuroQol Group has a robust tradition of encouraging the free exchange of ideas and material amongst its members. It has never been the intention or practice to suppress well-informed criticism - quite the reverse. Group members are actively encouraged to confront the shortcomings of their own science. Hence it is possible to find those within the Group who advocate for, say, an increased number of response categories from 3 to 5, or that additional dimensions should be added to the standard 5. The Group has historically determined its position on such matters only after deliberating on the empirical evidence with which it is presented. Someone must generate that evidence and in so doing they need to break away from previously held positions, or the very least to test their robustness. Such incremental decision-making has only been only sustainable given the high quality research and the exercise of self-discipline that has been the hallmark of the Group for so long.

Despite the high standards aspired to by the EuroQol Group, there remain some aspects of its evolution that have proved to be an irritant in later years - starting with

---

the name of the group itself. The founder members came from 4 countries (Finland, Netherlands, Sweden and the UK). They were concerned with measuring the value of health outcomes and since quality of life is heavily embedded in economic evalua-
tion, it seemed natural at the time to refer in an almost casual way to the research net-
work as the EuroQol Group. Continued usage led to the de facto labeling of that fraternity in precisely these terms. The term "EuroQol" for a while also (and unfortu-
nately) became synonymous with the instrument designed to capture self-rated health
status. Naming the component elements of what is now definitively termed EQ-5D
was later the subject of some further difficult internal discussions. Had its founders
only had the foresight to define and label in a standardized format then much of this
could have been avoided. Thus the reader here will need to be aware that different
authors may refer to the same components but using different terms. From an edito-
rial perspective it might have been preferable to substitute EQ-5D for "The EuroQol
Instrument" in the papers that form this volume. However, the intention in presenting
this collection in their original format is to try and preserve the essential elements of
the work using the language that was current within the Group at the time of writing.

The papers presented here span a 10-year period. They represent perhaps less than
10% of the total volume generated within the Group over that period and certainly
less than 5% of its total output. The collection bears witness to the breadth of the
developmental agenda that it tackled. They are rough diamonds, the unpolished
source material that fuelled the evolutionary process. The papers map the develop-
mental pathway that others have followed in pursuit of similar objectives. This collec-
tion specifically excludes more recent material since it was intended as a means of
charting the Group's initial research product. From its formative steps, the EuroQol
Group has evolved into a mature organization that continues to pursue its original
research objective of investigating the valuation of health but combines this with a
new role, as the corporate parent of EQ-5D. Were this volume to take account of
more recent material, it could report that there are now more than 80 different lan-
guage versions of EQ-5D and that soon a total of 2,000 studies will have been regis-
tered at the Group's website. The original group drew its members from amongst the
research community in a handful of countries. Today that membership is global. The
"Euro" in EuroQol was once mistakenly interpreted as indicating that EQ-5D was
designed for local European users. Today its user base is worldwide. For those users
who seek greater insights to the research work that underpins that status, this volume
provides some documentary evidence.

The process of assembling such a volume was always going to be a challenge. The
Group has produced a vast quantity of working papers. Its annual Scientific Plenary
meetings now involve discussion of some 30 new texts each year. In the early days
these meetings were twice as frequent but contained perhaps half the number of
papers. Over the decade to 1997 there were literally hundreds of papers from which
this selection was ultimately made. The cataloguing of this output has involved vari-
uous individuals, including Jan van Busschbach and Frank de Charro, given additional support by Erik Nord through the EuroQoLus system that he devised. Much of the early material was never archived in electronic form and the physical preparation of these papers required significant input from Kerry Atkinson and Ben Kind in York. Further refinement and all pre-production was carried out in Rotterdam by Dennis Kennedy working with Rosalind Rabin. Richard Brooks has once again reentered the fray in his editorial role. Their collective input to the creation of this volume has been indispensable.

And finally, acknowledgement is due to the one individual in the Group who provided the inspiration and leadership in its formative days, wise counsel and sagacity in its adolescent years and who continues to stimulate new activity in what is now the approach of its 20th anniversary. To Alan Williams.

Paul Kind
Centre for Health Economics, University of York
April 2005
List of tables and appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>2.1</td>
<td>Original 6D EuroQol descriptive system</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>2.2</td>
<td>5D EuroQol descriptive system</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>3.1</td>
<td>5-dimensional EuroQol questionnaire</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>5.1</td>
<td>Comparison of the index change before and after operation, for SIP, Rosser and EuroQol</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>5.2</td>
<td>Changes in health states per individual before and after hip joint surgery measured by EuroQol, SIP, and the Rosser Index</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>5.3</td>
<td>Average loss of health in percentage of full health for inpatient and out-patient casualties at Lund Hospital one week, one month, and six months after the accident, measured by three indexes: EuroQol, Rosser Index, and Thermometer</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>5.4</td>
<td>Weightings according to the Rosser Index</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>6.1</td>
<td>Layout of the English and Spanish EuroQol</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>6.2</td>
<td>Self-rated overall health status by socio-demographic variables</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>6.3</td>
<td>Self-rated overall health status by health-related variables</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>6.4</td>
<td>Average scores for the EuroQol Spanish study (n = 600)</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>6.5</td>
<td>Ratings of health states by self-rated overall health</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>6.6</td>
<td>Ratings of health states by age</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>6.7</td>
<td>Ratings of health states by level of education</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>6.8</td>
<td>Ratings of health states by degree of difficulty of the task</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>6.9</td>
<td>Ratings of health states by respondent’s health state</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>7.1</td>
<td>Dimensions, levels and the codes of the health states</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>7.2</td>
<td>The means and standard deviations of the health states</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>7.3</td>
<td>The actual time interval subjects used</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>7.4</td>
<td>The observed order of the valuations on pages 5 and 6</td>
<td>87</td>
</tr>
</tbody>
</table>
Table 7.5 Responses to the valuation of death on the EuroQol visual analog scale

Table 8.1 The EuroQol descriptive system (version 2)

Table 8.2 Demographic characteristics of the patients as reported by themselves

Table 8.3 Results of the EuroQol scaling task. For an explanation of health state notation, see Introduction and Table 8.1. The numbers in the table represent the patient’s self-rating of numerous hypothetical health states on a visual analogue scale from zero to 100. Patients were asked to value health states 11111 and 33333 twice to assess consistency and stability

Table 8.4 Valuations for the state ‘being dead’ in the EuroQol scaling task. Patients were asked to value this state (on a visual analogue scale from zero to 100) on 2 separate pages of the questionnaire, to allow assessment of consistency and stability

Table 8.5 Results of 3 pilot studies conducted with the EuroQol Instrument in 3 centres: (i) Lund, Sweden; (ii) Frome, England; and (iii) Bergen op Zoom, The Netherlands (BoZ) (EuroQol Group 1990, with permission[1]). For an explanation of health state notation, see 8.1 Summary and Table 8.1. The data in the table are the mean valuations given by participants for each health state, ranging from zero (worst) to 100 (best)

Table 9.1 Scheme for analysing test-retest reliability of health state valuations

Table 9.2 Response per version of the questionnaire

Table 9.3 Relevant background characteristics of respondents: total sample (only first survey) and test-retest sample (first and second survey)

Table 9.4 Valuations in test and retest (n = 208)

Table 9.5 Results of the Generalizability Study per version of the questionnaire

Table 10.1 Self-ratings and hypothetical valuations in the Finnish EuroQol study
List of tables and appendices

Table 10.2 Regression of hypothetical valuations and self-ratings on health dimensions 130

Table 10.3 Self-ratings in the Catalan EuroQol study 131

Table 10.4 Self-ratings and hypothetical valuations in the Catalan EuroQol study 132

Appendix 10.1 Studies of quality of life in patients 136

Table 11.1 The variants of the EuroQol questionnaire 140

Table 11.2 Percentage inconsistency rates for the lay concepts and Frome IV studies 142

Table 11.3 The effect of age on inconsistency rates 144

Table 11.4 The effect of education on inconsistency rates 144

Table 12.1 The trade-offs, computed utilities (computation different for health states valued worse than dead than for those health states which are valued better than dead) and the values after the York transformation for the TTO method 152

Table 13.1 Linear model, unrestricted, $x_{2i} = 0.5$, original values 164

Table 13.2 Linear model, unrestricted, $x_{2i} = \beta_i$, original values 164

Table 13.3 Linear model, restricted, $x_{2i} = 0.5$, rescaled values 165

Table 13.4 Linear model, restricted, $x_{2i} = \beta_i$, rescaled values 165

Table 13.5 Linear model, restricted, all individual values rescaled 166

Table 13.6 Multiplicative model, one first order effect, $x_{2i} = 0.5$, observed values 167

Table 13.7 Multiplicative model, first order effects, stepwise linear regression including main effects 168

Table 13.8 Multiplicative model, first order effects, stepwise linear regression 169

Table 14.1 Stem leave plot of the observations for state 33333 (TT033333 Stem Leaved) 173

Table 15.1 Mean, standard deviation and P-value for evaluation of the health states and ‘being dead’ by the 141 group and the 208 group 183
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 15.2</td>
<td>P-value for the comparison of ‘being dead’ and age</td>
<td>184</td>
</tr>
<tr>
<td>Table 15.3</td>
<td>P-value for the comparison of ‘being dead’ and sex</td>
<td>185</td>
</tr>
<tr>
<td>Table 15.4</td>
<td>P-value for the comparison of ‘being dead’ and education</td>
<td>185</td>
</tr>
<tr>
<td>Table 15.5</td>
<td>P-value for the comparison of ‘being dead’ and worked in health/social service</td>
<td>185</td>
</tr>
<tr>
<td>Table 15.6</td>
<td>P-value for the comparison of the evaluation of the health states and ‘being dead’</td>
<td>186</td>
</tr>
<tr>
<td>Table 16.1</td>
<td>Response rates, and the rate of usable responses in the sub-samples (N = 230 each)</td>
<td>192</td>
</tr>
<tr>
<td>Table 16.2</td>
<td>The distribution of respondents by age groups in each sub-sample before and after rejection of inconsistent responses and the percentage of the rejected responses</td>
<td>192</td>
</tr>
<tr>
<td>Table 16.3</td>
<td>The mean and median values of different health states in the Finnish EuroQol survey when their duration is defined to be 1 year, 10 years or unspecified</td>
<td>193</td>
</tr>
<tr>
<td>Table 16.4</td>
<td>Mean values and 95% confidence intervals in the convenience sample where the same respondents valued health states during 1 year, 10 years and 1 month time periods (N = 60)</td>
<td>195</td>
</tr>
<tr>
<td>Appendix 16.1</td>
<td>The standard EuroQol questionnaire pages 4-5-6</td>
<td>197</td>
</tr>
<tr>
<td>Table 17.1</td>
<td>Implied preferences from TTO valuations</td>
<td>206</td>
</tr>
<tr>
<td>Table 17.2</td>
<td>Preferences extracted from ranking task</td>
<td>208</td>
</tr>
<tr>
<td>Table 17.3</td>
<td>Upper segment of probability matrix corresponding to F-matrix based on TTO valuations</td>
<td>210</td>
</tr>
<tr>
<td>Table 17.4</td>
<td>Upper segment of probability matrix corresponding to F-matrix based on ranking task</td>
<td>211</td>
</tr>
<tr>
<td>Table 17.5</td>
<td>Scale values computed from implied TTO preferences</td>
<td>212</td>
</tr>
<tr>
<td>Table 17.6</td>
<td>Computed scale values from ranking task</td>
<td>213</td>
</tr>
<tr>
<td>Table 17.7</td>
<td>Decrements for pairwise model tariff and TTO Al tariff</td>
<td>215</td>
</tr>
<tr>
<td>Table 17.8</td>
<td>Tariff of values based on TTO preference matrix</td>
<td>217</td>
</tr>
<tr>
<td>Table 17.9</td>
<td>Tariff of values derived from ranking preferences</td>
<td>218</td>
</tr>
<tr>
<td>Table 18.1</td>
<td>Number of studies using the EQ-5D defined by clinical area</td>
<td>223</td>
</tr>
</tbody>
</table>
List of tables and appendices

Table 18.2 Number of studies using the EQ-5D identified within different countries 224

Table 18.3 Number of studies incorporating the EQ-5D defined by study design 225

Table 18.4 Sources of funding for studies incorporating the EQ-5D 225

Table 18.5 Generic instruments used alongside the EQ-5D (most studies incorporate more than one instrument) 226

Appendix 18.1 Three page form 227

Appendix 18.2 Titles of elicited studies as defined by clinical area 230
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Average loss of health during the first six months after the accident for light and moderate out-patient casualties measured by EuroQol and the Rosser index</td>
<td>58</td>
</tr>
<tr>
<td>7.1</td>
<td>EuroQol health state 33321</td>
<td>83</td>
</tr>
<tr>
<td>9.1</td>
<td>Page of the questionnaire</td>
<td>112</td>
</tr>
<tr>
<td>9.2</td>
<td>Values (z-scores) of health state valuations (first moment of measurement of version AB; n=52): Observed mean valuations and MDU scaled valuations, assuming respectively ordinal (MDU-ordinal) and interval data (MDU-interval)</td>
<td>118</td>
</tr>
<tr>
<td>12.1</td>
<td>Testing the effect of forcing dead to zero (complete transformation) for the VAS method on the data of the HESTEM experiments (Rotterdam)</td>
<td>150</td>
</tr>
<tr>
<td>12.2</td>
<td>Example of the situation for the time trade-off elicitation method of valuing a worse-than-dead health state two times worse (trade-off value = -5 years; utility = -1) than the (non) health state ‘dead’</td>
<td>152</td>
</tr>
<tr>
<td>12.3</td>
<td>The effect on the distribution of backward transformation for valuations valued worse than dead by the time trade-off method</td>
<td>154</td>
</tr>
<tr>
<td>13.1</td>
<td>Health state values from the EuroQol Rotterdam survey</td>
<td>163</td>
</tr>
<tr>
<td>14.1</td>
<td>Median and mean TTO data</td>
<td>172</td>
</tr>
<tr>
<td>14.2</td>
<td>TTO33333 scattered</td>
<td>173</td>
</tr>
<tr>
<td>14.3</td>
<td>Plus transformation</td>
<td>175</td>
</tr>
<tr>
<td>14.4</td>
<td>Two medians</td>
<td>177</td>
</tr>
<tr>
<td>15.1</td>
<td>The first sheet of the EuroQol valuation exercise</td>
<td>182</td>
</tr>
<tr>
<td>15.2</td>
<td>The evaluation of health states</td>
<td>184</td>
</tr>
<tr>
<td>17.1</td>
<td>Graphical representation of classical Thurstone model</td>
<td>202</td>
</tr>
<tr>
<td>17.2</td>
<td>Analysis of preferences in TTO data</td>
<td>214</td>
</tr>
<tr>
<td>17.3</td>
<td>Analysis of preferences in ranking data</td>
<td>214</td>
</tr>
</tbody>
</table>
List of figures

**Figure 17.4** Estimated values for directly observed health states based on TTO and ranking preferences  216

**Figure 17.5** Tariff values estimated from the observed TTO data and corresponding values in a tariff estimated from the ranking task  216

**Figure 18.1** Results from an international survey showing 19 clinical areas covered by the EQ-5D  222

**Figure 18.2** Results from an international survey showing areas where studies using the EQ-5D are being undertaken  224

**Figure 18.3** Results from an international survey showing types of studies using the EQ-5D  225
The raison d’être of the EuroQol Instrument is to provide a simple “abstracting” device, for use alongside other more detailed measures of health-related quality of life (henceforth HRQoL), to serve as a basis for comparing health care outcomes using a basic “common core” of QoL characteristics which most people are known to value highly. From the outset it was accepted that for such comparisons to be useful it would be necessary to go beyond generating such information in the form of a “profile” (though the EuroQol data can be used in that way too), and therefore the issue of the relative valuation of different health states had to be confronted. It was further recognised that such information would be extremely valuable in a QALY-type context, but for that purpose it would be necessary to include a valuation for the state of being dead, otherwise it would be impossible to establish a cardinal index scale in which “dead” = 0 and “healthy” = 1, a property that is required for QALY-type calculations.

1.1 RESEARCH STRATEGY

People value both improvements in life expectancy and improvements in the quality of their lives, therefore any single index of the benefits of health care has to incorporate both. Ideally we would like to know how every individual values every possible prospective time profile of HRQoL, including the probabilities associated with each component in them. It need hardly be said that such an undertaking is impossible, so some strategic decisions have to be made concerning simplification of this research task. It is impossible to tell a priori which simplifications are best, so we are inevitably in the realm of intuition and scholarly judgement. It would therefore be most unwise for the research community to pursue only one strategy. What is called for is a broad range of different approaches, which need to be periodically reviewed to see what seems to be working and what does not.

1.2 THE DESCRIPTION OF HEALTH PROSPECTS

The particular simplifications involved in the approach adopted by the EuroQol Group were as follows:

(i) Each time profile of prospective health states is divided into separate time segments, such that within each segment the HRQoL of the individual is constant.

(ii) Initially it was to be assumed that the relative values attached to the different states were independent of the states that preceded or succeeded them, and of the length of time spent in each state.
Thus the Group rejected the scenario approach to health state description, in favour of using composite health states set out in a standardised manner. It was anticipated that at a later stage assumption (ii) could be relaxed, but in the meantime a common time duration was to be used (1 year), with the subject told that what happened thereafter was not known and should not be taken into account. The relaxation of this assumption is now under way within the Group, with data now becoming available from surveys undertaken in Finland and in The Netherlands. We too have experimental work about to start in the UK on the effect of varying the duration and the sequencing of states. Preliminary results are expected in October 1993.

A second set of simplifications is required concerning the detailed description of the health states themselves. There is an understandable tendency in this kind of enterprise to include everything that might be of any interest to anyone, and to work with fine enough gradations of “severity” within each “dimension” of HRQoL to pick up any effects of health care treatment that might be of interest to a discriminating practitioner. It is important here to recall that a 10-dimensional classification, with ten levels within each dimension, yields a classification system with $10^{10}$, or 10 000 million different cells. Such detail is quite inappropriate in an abstracting device, and if there is to be a reasonable prospect that such summary data is to be collected readily, it must be very simple to collect. If, for some other purpose, more detailed data is required it should be collected by an instrument designed for that specific purpose. Thus the general advice offered to prospective users is to use the (very simple) EuroQol Instrument alongside (not instead of) a more detailed specific measure, and, at this developmental stage in HRQoL measurement, preferably also alongside some more comprehensive generic measure (possibly one using the profile approach). This is now generating “calibration” data enabling systematic comparisons to be made between outcomes as measured by the EuroQol Instrument and outcomes as measured by other instruments.

The actual choice of descriptive content in the EuroQol Instrument originated from a review of existing instruments, the material so culled being reduced to manageable proportions through discussion between the original members of the Group, who ranged across many disciplines and who drew collectively on a wide range of experience. Later there was an opportunity to test these judgements against the results of a survey of lay concepts of health, which suggested that a dimension of “energy/tiredness” ought to be added to the original 6 dimensions. To accommodate this, 2 existing dimensions (1 concerned with work activities and the other with other activities) were fused, since it had transpired that the “other activities” dimension added little to the overall valuations of states. It turned out, however, that “energy/tiredness” also made little contribution to health state valuations, and in the pursuit of parsimony it was therefore dropped. This left us with the present 5-dimensional set, in which each dimension has 3 levels of “severity”, generating 243 different cells (see Appendix 1.1). To these need to be added “unconscious” (because it cannot be regarded as a
“composite” of the 5 dimensions) and “dead” (because it is required as a calibration point on the 0 to 1 scale), making 245 states in all.

1.3 THE VALUATION OF HEALTH PROSPECTS

Whilst in the process of establishing a workable descriptive system, the Group had also been devoting a great deal of attention to valuation issues. Early on it had been agreed that relative valuations should be sought for composite (multi-dimensional) states, not for each dimension separately. This important decision complicated the valuation task, because it involved rejection of the multi-attribute utility scaling approach, in case there proved to be significant interaction between the dimensions. But since no one subject could be expected to value more than a dozen or so states, this meant that the choice of states to be valued had to be made in such a way that, if necessary, it would be possible to estimate the values of all the other states from that limited number of observations. Thus a standard minimum set of (14) states was chosen which were to be used in all valuation work by all members, though where possible members were encouraged also to elicit values for a more extended set. From these states (plus a value for “being dead”) the whole valuation space needs to be estimated. This is a task with which we are still experimenting with a variety of estimation techniques to see which uses our data most fully and produces the best fit.

For practical reasons the EuroQol Group imposed upon itself a very restrictive condition concerning the main body of data collection on relative valuations, namely that the questionnaire design should be so simple that it could be self-completed and conducted by a postal survey. This was essentially because none of us had research funds sizeable enough for any other alternative to be feasible, given that we wanted valuations from a general public, not from convenience samples. We quickly agreed that the only valuation method that would be practicable in that context was the visual analogue scale (VAS), and for this purpose we adopted a thermometer-like scale, the current form of which (see Appendix 1.1) is the result of considerable experimentation (using shorter, longer, differently calibrated, differently labelled, and differently orientated versions). The associated problem, which was also subjected to a fair amount of empirical testing, was the layout within the questionnaire of the states to be valued. A complication here was that in order to standardise “framing” effects we had to repeat some states on subsequent pages, thereby reducing the number of observations available to us. But the most difficult issue with the visual analogue scale has been getting people to value the state of “being dead” alongside the other states, and this is still an active area of experimentation within the Group.

Lately, with more research funds being devoted to this kind of research, it has been possible to generate valuations for EuroQol states using valuation methods other than the VAS, and in particular comparing those valuations with ones derived from the Standard-Gamble (SG) and the Time-Trade-Off (TTO) methods. From preliminary
work, there appears to be a power-function relationship between the VAS and each of the other 2 methods. It further appears that the TTO method yields somewhat better quality data than the SG method, if quality is judged by the internal consistency of the answers given by respondents, the sensitivity of valuations to parameters known to influence them, and the reliability of the responses when the valuation task is repeated by the same respondents some weeks later. For this reason the TTO method will be used alongside the VAS (thermometer) in the next round of our own work, which is now under way.

This next round of work is to consist of just over 3000 interviews with a representative sample of the adult population of the UK living in their own homes. Valuations will be sought on about 40 EuroQol states, carefully selected so as to be well spread through the valuation space, and to be particularly useful for the estimation of the values of the states on which we shall have no direct valuations. Respondents will each value a stratified random sample of 15 of these 40 or so states, using first a simple ranking of states, then a rating on the “thermometer” (using the method of “bisection”, so as to generate an interval scale), and finally the TTO method. The fieldwork is being conducted with Social and Community Planning and Research (SCPR), who have worked with us over the past year or so to develop this interview-based format. We anticipate that the fieldwork will be completed by the end of 1993, and the preliminary results available around Easter 1994. We hope to be able to deliver a full report to the Health Economics Study Group (HESG) at the Summer Meeting in 1994. At that point we expect to have a tariff of values for all EuroQol states which will be representative of the views of the UK public. We are at the very early stages of developing a parallel study to elicit the values of doctors and nurses by identical methods, to see how they compare with each other, and with the views of the general public.

A widespread concern with all HRQoL measurement is the validity and reliability of each particular measure. In general, establishing “validity” requires the investigator to address the question “does your measure measure what it purports to measure?”. But since there is no “gold standard” for the measurement of health-related quality of life, this seems an unanswerable question. So what people fall back on instead are appeals to plausibility, for instance: testing whether the measure contains the kind of elements that we would expect such a measure to have; whether it goes up when we would expect it to go up and down when we would expect it to go down; and so on. These are all very subjective notions, and ultimately rely heavily on intuition and professional judgement. We know from our own earlier work:
(i) that EuroQol self-rated health on the VAS thermometer declines (as expected) with age, and

(ii) that whilst self-reported pain and discomfort increase with age, anxiety and depression decrease with age, and

(iii) that people’s valuations are affected both by age and by experience of illness.

My own personal view is that searching for “validity” in this field, at this stage in the history of HRQoL measurement, is like chasing a will o’ the wisp, and probably equally unproductive. It would, however, be useful to find out whether the values elicited from particular individuals in particular circumstances are consistent with their actual behaviour when they are put in a situation in which those values should have been crucial. The devising, conduct and interpretation of the results of such a study would be a valuable contribution to this otherwise rather murky area.

Reliability refers to the issue as to whether the values elicited from an individual are stable, which is usually tested by (surreptitiously?) repeating a question at different stages within an interview, or by going back to the individual a short while later in order to see whether on the second occasion the same answers are obtained as on the first occasion. Within the context of a short self-completed questionnaire the former method does not seem appropriate, and if (as they usually are) questionnaires are returned anonymously, the latter method is not possible. But in recent interviewer-led survey work we have been able to use the latter method, and it appears that the valuations elicited in that context are stable and reliable.

1.4 THE EUROQOL INSTRUMENT AND THE MEASUREMENT OF QUALITY OF LIFE

The EuroQol Instrument has two distinct contributions to make to the task of measuring health-related quality of life. First, it offers a very convenient way of collecting descriptive data about HRQoL, and about people’s own self-rating of their current health state (by using only pages 2 and 3 of the questionnaire in Appendix 1.1). This descriptive data is needed if we are to fill the gaps in our knowledge about the HRQoL sequelae of many common health care activities. The second, and much more ambitious, role is that of supplying a tariff of social values of health states, to be used (alongside cost data) in a planning context when determining priorities for health care. Each of these distinct, but related, roles will now be considered in turn.

1.5 THE DESCRIPTIVE ROLE

Concerning the descriptive role of the data which can be collected using only pages 2 and 3 of the EuroQol questionnaire, this produces (from page 2) a simple description by patients of their health-related quality of life. This requires only that five ticks are entered on to the form, which is essentially just page 2 of Appendix 1.1.
It takes less than a minute to complete, and can either be self-assessed or observer-assessed. The data can be summarised, in profile form, as a simple 5-digit code. There are, however, considerable advantages in going beyond this stage and adding the self-rating exercise shown on page 3 of Appendix 1.1. These additional data enable a single summary statistic to be generated, which can either be used in conjunction with the 5-digit code, or as a measure of self-rated health in its own right. It may be analysed either in a cross-sectional manner when pooled with similar data from other subjects, or it may be used in a longitudinal manner to trace out each individual’s rating of own health at different points in time. It is probably more safely used in the latter manner than in the former, since optimistic people may rate all states higher than pessimistic people do, but each will rate the direction of change in their health accurately.

As was indicated earlier, the EuroQol Instrument was not designed as a “stand alone” instrument for measuring all kinds of HRQoL in sufficient detail for all purposes. It was designed as an “abstracting” device, to be used alongside other measures as necessary, and intended to provide information about a common core of key items, which should always be of interest, because they represent people’s salient concerns about HQQoL. Because of its summary, generic nature, it is likely that it will normally be used alongside other more detailed measures, either of a generic or of a specific nature, which focus on the particular concerns of the investigators in each study.

This opens up the possibility of systematic within-subject comparison of the EuroQol descriptions with the descriptions generated by other methods. This would enable these other descriptions to be given a EuroQol score, using our tariff. Where the other methods also generate a single summary score, it also opens up the possibility of recalibrating these other scores using the EuroQol valuations. Once sufficient data of this comparative kind were available, it should be possible to attach EuroQol scores directly to data which was collected solely with other instruments. Where the alternative instrument is a generic instrument (such as the Nottingham Health Profile (NHP) or Short Form 36 (SF-36)) it should be possible to establish a fairly close mapping of the more detailed instrument onto the simpler EuroQol system. It may be more difficult with disease-specific or treatment-specific measures, since they will typically not cover such a broad range of dimensions as a generic measure, concentrating on depth rather than breadth of coverage.

There are, therefore, limitations to this enterprise. One is that, even where a broad range of HRQoL dimensions is tracked by other instruments, they are not usually reported in sufficient detail for this recalibration to be possible. To move from a series of frequency distributions relating to a group of patients in a trial, to the particular combination of characteristics manifested by individual subjects, requires quite sweeping assumptions about the interrelationships between the different elements which may be difficult to sustain. This means that it may be necessary to engage in