OUTCOMES OF
ORTHOGNATHIC TREATMENT

SUSAN JANE CUNNINGHAM
BChD (Hons); FDSRCS (Orth); MSc (London);
MOOrth (RCS England)
2000

Department of Orthodontics, Eastman Dental Institute for Oral
Health Care Sciences, University College London

This report is submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy
ABSTRACT

In recent years health care evaluation has become consumer based. The main source of new information regarding health outcomes is the standardised patient survey, for example: generic measures of well-being. To date, there is little research looking at outcomes of orthognathic treatment in this way.

The first part of this research (Chapter 2) studied the economic evaluation of orthognathic treatment, in particular cost-utility analysis and willingness-to-pay (WTP). Utility values obtained from patients and controls were found to be comparable for two of the three methods used and confirmed the view that non-patients may provide accurate values if the health state in question is described well. A longitudinal study of patients allowed calculation of quality adjusted life years (QALYs) gained as a result of treatment, with mean non-discounted values of between 8.13 and 14.02, depending on the methodology used. The mean cost per QALY varied from a minimum value of £205.79 to a maximum of £1,073.78. The concept of WTP was also introduced and went some way to confirming that it measures strength of preference.

The second part of the study (Chapter 3) aimed to develop a condition-specific quality of life measure for patients requesting orthognathic treatment. Evidence was presented for the reliability, validity and responsiveness of the questionnaire. Significant improvements in quality of life were noted following treatment.

The final section (Chapter 4) used the relatively new method of statistical modelling to study changes in a number of psychometric instruments during treatment. There were differences in psychological profile between patients and controls at the start of treatment and significant positive changes were noted following intervention.

In conclusion, all three chapters found that orthognathic treatment was associated with positive outcomes. Research methodology of this type is increasingly important in dentistry and larger scale multi-centre trials are recommended for the future.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Title</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Contents</td>
<td>3</td>
</tr>
<tr>
<td>List of tables</td>
<td>7</td>
</tr>
<tr>
<td>List of figures</td>
<td>11</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>13</td>
</tr>
<tr>
<td>List of publications</td>
<td>15</td>
</tr>
</tbody>
</table>

## CHAPTER 1

### INTRODUCTION

1.0 Review of the literature 20
   1.1 Introduction 20
   1.2 Clinical outcome measures 23
   1.3 Economic outcome measures 24
   1.4 Personal outcome measures 25
   1.5 Psychometric properties of instruments 28

2.0 Summary 33

## CHAPTER 2

### ECONOMIC EVALUATION OF ORTHOGNATHIC TREATMENT 34

1.0 Review of the literature
   1.1 Introduction 35
   1.2 Allocation of resources 36
   1.3 Costing health care 38
   1.4 Methods of economic evaluation 43
   1.5 Utility analysis 46
   1.6 The cost-benefit approach 64
   1.7 Economic evaluation in dentistry 73
1.8 Summary .................................................. 79

2.0 Aims ......................................................... 80

3.0 Materials and methods ..................................... 81
   3.1 Instruments .............................................. 81
   3.2 Subjects ................................................ 85
   3.3 Repeatability of the method ................................ 89
   3.4 Calculation of QALYs gained ........................... 89
   3.5 Estimation of costs ...................................... 93
   3.6 Sensitivity analyses .................................... 93
   3.7 Discounting costs and benefits ......................... 96
   3.8 Bootstrapping of results ............................... 97
   3.9 Statistical software and procedures ................... 98

4.0 Results: Utility analysis ................................... 99

5.0 Results: Willingness-to-pay ............................... 145

6.0 Discussion ............................................... 148
   6.1 General discussion ..................................... 148
   6.2 Comparison of pre-treatment utility values from both experimental and control groups .......... 151
   6.3 Longitudinal utility analysis .......................... 155
   6.4 Willingness-to-pay ..................................... 165

7.0 Conclusions .............................................. 167

8.0 Suggestions for further study ............................. 169

CHAPTER 3
QUALITY OF LIFE IN ORTHOGNATHIC PATIENTS ............... 170

1.0 Review of the literature .................................. 170
   1.1 Introduction ........................................... 171
   1.2 Quality of life models ................................ 173
   1.3 Quality of life and health care outcomes ............. 175
   1.4 Methodological issues ................................. 176
CONTENTS

1.5 Multilevel modelling 253

2.0 Aims 257

3.0 Materials and methods

3.1 Contents of the questionnaire 258

3.2 Reliability 260

3.3 Subjects 260

3.4 Statistical analysis 265

3.5 Studies undertaken 273

4.0 Results 274

5.0 Discussion 306

5.1 Reliability and validity 306

5.2 Use of multilevel multivariate analysis 306

5.3 Study 1: Comparison of experimental and control groups 307

5.4 Study 2: Does pre-surgical orthodontics result in adverse psychological effects? 314

5.5 Study 3: Longitudinal data analyses 318

5.6 Overall summary of findings 325

6.0 Conclusions 329

7.0 Suggestions for further study 332

GENERAL CONCLUSIONS 333

APPENDICES 336

Appendix 1 Details of RS, SG and TTO 337

Appendix 2 Examples of individual utility values 342

Appendix 3 Resource use data 346

Appendix 3 Condition-specific quality of life measure (OQLQ) 364

Appendix 4 Short Form-36 368

Appendix 5 Psychometric measures used in Chapter 4 377

REFERENCES 390
LIST OF TABLES

CHAPTER 1

Table 1  Benefits of orthognathic treatment  22
Table 2  Risks of orthognathic treatment  22
Table 3  Uses of patient rated oral health indicators  33

CHAPTER 2

Table 4  Utility values for different health states  56
Table 5  Example of a QALY league table  62
Table 6  Utility values for chronic health conditions  64
Table 7  Sample size calculations (experimental/control)  86
Table 8  Sample size calculations (longitudinal study)  87
Table 9  Details of respondents  99
Table 10  Test-retest data  100
Table 11  Comparison of utility values for both groups using all 3 methods  101
Table 12  Comparison of mean utility values between the experimental and control groups  102
Table 13  Spearman correlations between age and utility values for the experimental and control groups  102
Table 14  Method agreement between the 3 utility measurement methods  103
Table 15  Details of respondents (longitudinal study)  104
Table 16  Utility values for all three methods at all five time periods  107
Table 17  Comparison of bimaxillary and single jaw values at T1 and T5  108
Table 18  T1/T5 changes  108
Table 19  Mean QALYs for all three methods (groups combined)  110
Table 20  Mean QALYs for all three methods (bimaxillary group)  111
Table 21  Mean QALYs for all three methods (single jaw group)  112
Table 22  Mean total QALYs  113
Table 23 Comparison of 95\% confidence intervals for total QALYs gained using conventional and bootstrap methods 116
Table 24 Sensitivity analysis of total QALYs gained 118
Table 25 A sensitivity analysis of orthodontic treatment costs 120
Table 26 A sensitivity analysis of in-patient hotel costs 123
Table 27 A sensitivity analysis of operating theatre costs 126
Table 28 A sensitivity analysis of total resource use 129
Table 29 Mean non-discounted and discounted total resource use 132
Table 30 Comparison of 95\% confidence intervals for total costs as derived by conventional and bootstrap methods 135
Table 31 Mean cost per QALY 137
Table 32 Mean cost per QALY with 95\% CI from bootstrap method 138
Table 33 Sensitivity analysis of cost per QALY (groups combined) 141
Table 34 WTP: Reliability test-retest data 145
Table 35 Willingness-to-pay values 145
Table 36 Comparison of WTP between the two groups 146
Table 37 Relationship between WTP and utility values and WTP and Townsend deprivation scores (experimental group) 146
Table 38 Correlations between WTP values and age 146
Table 39 Change in WTP in the longitudinal group (T1/T5) 147

CHAPTER 3

Table 40 Comparison of different methods of HRQL instrument administration 179
Table 41 Comparison of evaluative and discriminative instruments 181
Table 42 Comparison of generic and condition-specific HRQL instruments 182
Table 43 Oral health-related quality of life instruments 191
Table 44 Item reduction 205
Table 45 Principal component analysis (PCA) data 208
### LIST OF TABLES

| Table 46 | The rotated factor matrix (PCA) | 209 |
| Table 47 | T1 data | 210 |
| Table 48 | Gender differences for domain scores | 211 |
| Table 49 | Reliability testing: internal consistency | 212 |
| Table 50 | Reliability testing: test-retest data | 213 |
| Table 51 | Validity testing: Spearman correlations | 215 |
| Table 52 | Validity testing: Comparison of hypothesised and actual findings | 216 |
| Table 53 | Validity testing: Comparison of hypothesised and actual findings | 217 |
| Table 54 | Data for the responsiveness testing group | 218 |
| Table 55 | Responsiveness testing: OQLQ and VAS | 219 |
| Table 56 | Responsiveness testing: SF-36 | 220 |
| Table 57 | SF-36 data for the responsiveness testing group | 222 |
| Table 58 | Comparison of SF-36 data with data from the Oxford study | 223 |
| Table 59 | Validity, reliability and responsiveness: How does the OQLQ compare? | 235 |

### CHAPTER 4

| Table 60 | Scoring of questionnaires | 259 |
| Table 61 | Sample size calculation (control vs experimental) | 261 |
| Table 62 | Sample size calculation (longitudinal group) | 263 |
| Table 63 | Reliability of the questionnaires | 276 |
| Table 64 | T1: Comparison of experimental and control group | 277 |
| Table 65 | T2: Respondent details | 277 |
| Table 66 | T3: Respondent details | 278 |
| Table 67 | Comparison of experimental and control group: values prior to transformation and standardisation (Study 1) | 280 |
| Table 68 | Comparison of experimental and control group: significant parameters for each outcome variable (Study 1) | 281 |
Table 69  Comparison of experimental and control group: significant coefficients for each outcome variable (Study 1) 282

Table 70  Comparison of experimental and control group: regression coefficients converted back to original outcome measure scales (Study 1) 283

Table 71  Comparison of experimental and control group: correlations for the null model (Study 1) 284

Table 72  T2/T1 comparison: Values for changes in T2/T1 scores 286

Table 73  T2/T1 comparison: significant parameters for each outcome variable 287

Table 74  T2/T1 comparison: significant coefficients for each outcome variable 288

Table 75  T2/T1 comparison: regression coefficients converted back to original outcome measure scales 289

Table 76  T2/T1 comparison: correlations for the null model 290

Table 77  Longitudinal analysis: p-values from Wilcoxon signed-rank test and MLM null model 295

Table 78  T2/T1 comparison: re-run in the same format as T3/T2 and T3/T1 296

Table 79  T3/T2 comparison: Values for changes in scores between T3/T2 297

Table 80  T3/T2 comparison: significant coefficients for each outcome variable 298

Table 81  T3/T2 comparison: regression coefficients converted back to original outcome measure scales 299

Table 82  T3/T1 comparison: Values for changes in scores between T3/T1 303

Table 83  T3/T1 comparison: significant coefficients for each outcome variable 304

Table 84  T3/T1 comparison: regression coefficients converted back to original outcome measure scales 305
# LIST OF FIGURES

## CHAPTER 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The standard gamble technique</td>
<td>49</td>
</tr>
<tr>
<td>2</td>
<td>The time trade-off technique</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>Effect of health care interventions on quality and quantity of life</td>
<td>59</td>
</tr>
<tr>
<td>4</td>
<td>The standard gamble technique</td>
<td>83</td>
</tr>
<tr>
<td>5</td>
<td>The time trade-off technique</td>
<td>84</td>
</tr>
<tr>
<td>6</td>
<td>QALYs lost/gained during treatment</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>Calculation of total QALYs gained</td>
<td>92</td>
</tr>
<tr>
<td>8</td>
<td>Summary plot for all 21 subjects (RS method)</td>
<td>105</td>
</tr>
<tr>
<td>9</td>
<td>Summary plot for all 21 subjects (SG method)</td>
<td>105</td>
</tr>
<tr>
<td>10</td>
<td>Summary plot for all 21 subjects (TTO method)</td>
<td>106</td>
</tr>
<tr>
<td>11</td>
<td>Mean non-discounted and discounted QALYs: TTO (both groups)</td>
<td>114</td>
</tr>
<tr>
<td>12</td>
<td>Mean non-discounted and discounted QALYs: TTO (bimaxillary)</td>
<td>114</td>
</tr>
<tr>
<td>13</td>
<td>Mean non-discounted and discounted QALYs: TTO (single jaw)</td>
<td>115</td>
</tr>
<tr>
<td>14</td>
<td>Sensitivity analysis of QALYs gained</td>
<td>117</td>
</tr>
<tr>
<td>15</td>
<td>Sensitivity analysis of orthodontic treatment costs (both groups)</td>
<td>121</td>
</tr>
<tr>
<td>16</td>
<td>Sensitivity analysis of orthodontic treatment costs (bimaxillary)</td>
<td>121</td>
</tr>
<tr>
<td>17</td>
<td>Sensitivity analysis of orthodontic treatment costs (single jaw)</td>
<td>122</td>
</tr>
<tr>
<td>18</td>
<td>Sensitivity analysis of in-patient hotel costs (both groups)</td>
<td>124</td>
</tr>
<tr>
<td>19</td>
<td>Sensitivity analysis of in-patient hotel costs (bimaxillary)</td>
<td>124</td>
</tr>
<tr>
<td>20</td>
<td>Sensitivity analysis of in-patient hotel costs (single jaw)</td>
<td>125</td>
</tr>
<tr>
<td>21</td>
<td>Sensitivity analysis of operating theatre costs (both groups)</td>
<td>127</td>
</tr>
<tr>
<td>22</td>
<td>Sensitivity analysis of operating theatre costs (bimaxillary)</td>
<td>127</td>
</tr>
<tr>
<td>23</td>
<td>Sensitivity analysis of operating theatre costs (single jaw)</td>
<td>128</td>
</tr>
<tr>
<td>24</td>
<td>Sensitivity analysis of total resource use (both groups)</td>
<td>130</td>
</tr>
<tr>
<td>25</td>
<td>Sensitivity analysis of total resource use (bimaxillary)</td>
<td>130</td>
</tr>
<tr>
<td>26</td>
<td>Sensitivity analysis of total resource (single jaw)</td>
<td>131</td>
</tr>
<tr>
<td>27</td>
<td>Mean non-discounted and discounted total resource use (both groups)</td>
<td>133</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>28</td>
<td>Mean non-discounted and discounted total resource use (bimax)</td>
<td>133</td>
</tr>
<tr>
<td>29</td>
<td>Mean non-discounted and discounted total resource use (single jaw)</td>
<td>134</td>
</tr>
<tr>
<td>30</td>
<td>Mean cost per QALY (non-discounted and discounted: base)</td>
<td>139</td>
</tr>
<tr>
<td>31</td>
<td>Mean cost per QALY (non-discounted and discounted: low)</td>
<td>139</td>
</tr>
<tr>
<td>32</td>
<td>Mean cost per QALY (non-discounted and discounted: high)</td>
<td>140</td>
</tr>
<tr>
<td>33</td>
<td>Sensitivity analysis of cost per QALY (non-discounted)</td>
<td>142</td>
</tr>
<tr>
<td>34</td>
<td>Sensitivity analysis of cost per QALY (discounted)</td>
<td>142</td>
</tr>
</tbody>
</table>

**CHAPTER 3**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>Responsiveness testing: comparison of mean OQLQ domain scores at all three time points</td>
<td>221</td>
</tr>
<tr>
<td>36</td>
<td>SF-36 data at T2 and T3</td>
<td>224</td>
</tr>
<tr>
<td>37</td>
<td>SF-36 data compared with data from the Oxford study</td>
<td>224</td>
</tr>
</tbody>
</table>

**CHAPTER 4**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>Example of a multilevel data structure in dentistry</td>
<td>255</td>
</tr>
<tr>
<td>39</td>
<td>Multilevel data structure</td>
<td>265</td>
</tr>
<tr>
<td>40</td>
<td>The relationship between patient-perceived severity and the body image index</td>
<td>291</td>
</tr>
<tr>
<td>41</td>
<td>Distribution of scores for the Body Image Index at T1</td>
<td>292</td>
</tr>
<tr>
<td>42</td>
<td>State anxiety: changes over the longitudinal periods T2/T1 and T3/T2 (females)</td>
<td>300</td>
</tr>
<tr>
<td>43</td>
<td>State anxiety: changes over the longitudinal periods T2/T1 and T3/T2 (males)</td>
<td>301</td>
</tr>
<tr>
<td>44</td>
<td>Trait anxiety: changes over the longitudinal periods T2/T1 and T3/T2</td>
<td>302</td>
</tr>
<tr>
<td>45</td>
<td>Interaction variable representation</td>
<td>323</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors, Professor Nigel Hunt, Dr Mark Gilthorpe and Dr Charlotte Feinmann, for their guidance and encouragement during this work. I am extremely grateful to them. I would also like to thank Professor Hunt for his support throughout my orthodontic career.

In addition, I am very grateful to the following people for their invaluable help:
Dr Franco Sassi (London School of Economics and Political Science) for his advice in setting-up the utility analysis study. Thank you also for encouraging me that this work was worth pursuing.

Professor George Torrance (McMaster University, Ontario) for finding the time in a busy schedule to offer his advice regarding measurement of utility values.

Dr Mark Sculpher (Centre for Health Economics, University of York) for his advice regarding costing and cost-utility analysis and also for his enthusiasm regarding this work.

Dr Andrew Garratt (Unit of Health Care Epidemiology, Institute of Health Sciences, University of Oxford) for his invaluable help with the quality of life research.

Dr Emma MacIntosh (Health Economics Research Unit, Department of Public Health, University of Aberdeen) for her advice regarding willingness-to-pay methodology.

Professor Stanton Newman (Department of Academic Psychiatry, University College London) for his advice regarding choice of psychometric tests.

Miss Julie Barber (MRC Biostatistics Unit, Cambridge) for her assistance with the bootstrapping technique in Chapter 2.
A major thanks goes again to Dr Mark Gilthorpe (Unit of Biostatistics, Eastman Dental Institute) for teaching me the intricacies of MLwiN and for being very patient during the early part of my learning curve! Also to Dr Aviva Petrie (Unit of Biostatistics, Eastman Dental Institute) for her statistics advice.

A big thank you to all the staff at the Eastman Dental Institute who have given me a great deal of encouragement over the last few years, particularly Sheelah Harrison for advice based on her own PhD.

Thank you to all the consultants who allowed me access to their patients and thank you to all the patients who gave so freely and uncomplainingly of their time, it was a pleasure to work with them. In addition, thanks to members of the general public who acted as control subjects.

Thanks to all my very supportive friends, I promise I will be more sociable now!

Finally, the most important thank you goes to my parents and my sister, Jill, for their support (both emotional and financial!) over my long years of training. Thank you for always being there, I would never have achieved this without you.
PAPERS ACCEPTED BY PEER REVIEWED JOURNALS

1. Cunningham S. J.
   Economic evaluation of health care - is it important to us?

2. Cunningham S. J., Gilthorpe, M. S. and Hunt N. P.
   Are orthognathic patients different?

3. Cunningham S. J., Garratt A. and Hunt N. P.
   Development of a condition-specific quality of life measure for patients with
dentofacial deformity: I. Reliability of the instrument.

4. Cunningham S. J. and Hunt N. P.
   Relationship between utility values and willingness to pay in patients undergoing
orthognathic treatment.
   Community Dental Health 2000; 17: 92-96.

5. Cunningham S. J. and Hunt N. P.
   A comparison of health state utilities for dentofacial deformity as derived from
patients and members of the general public.

6. Gilthorpe M. S. and Cunningham S. J.
   The application of multilevel modelling to multivariate dental research data.
   Community Dental Health
   Accepted September 1999. In press.
7. **Cunningham S.J.** and Hunt N.P.
   Quality of life and its importance in orthodontics.
   Journal of Orthodontics.
   Accepted August 2000. In press.

PAPERS SUBMITTED TO PEER REVIEWED JOURNALS

1. **Cunningham S.J.**, Garratt A. and Hunt N.P.
   Development of a condition-specific quality of life measure for patients with
dentofacial deformity: II. Validity and responsiveness testing.
   Community Dentistry and Oral Epidemiology.
   Submitted March 2000 (Revisions submitted July 2000).

PAPERS IN PREPARATION

1. **Cunningham S.J.**, Gilthorpe M.S. and Hunt N.P.
   Does pre-surgical orthodontics result in adverse psychological effects?
   Due to be submitted August 2000.

2. **Cunningham S.J.**, Sculpher M. and Sassi F.
   Due to be submitted September/October 2000.
REFEREED ABSTRACTS

1. Cunningham S.J., Gilthorpe M.S. and Hunt N.P.
   Are orthognathic patients psychologically different from a control group?

POSTER PRESENTATIONS

1. Cunningham S.J., Gilthorpe M.S. and Hunt N.P.
   Are orthognathic patients psychologically different from a control group?
   European Orthodontic Conference, Strasbourg, June 1999
   and
   British Orthodontic Conference, Glasgow, September 1999.

2. Cunningham S.J. and Hunt N.P.
   A utility analysis of patients requesting orthognathic treatment for the correction of dentofacial deformity.
   and
   British Orthodontic Conference, Glasgow, September 1999.
Dedicated to my mum and dad
CHAPTER 1

OUTCOMES OF ORTHOGNATHIC TREATMENT: AN INTRODUCTION
1.0 REVIEW OF THE LITERATURE

1.1 INTRODUCTION

Recent years have seen numerous changes in which health care evaluation has become consumer based and both generic and specific outcomes are increasingly important. Purchasers, clinicians, managers and policy analysts all use information about well-being, functional status and satisfaction with treatment. The main source of new information regarding health care and health outcomes is rapidly becoming the standardised patient survey investigating areas such as: effects of specific diseases or treatments; generic measures of physical and mental well-being; condition-specific measures; evaluation of specific features of health care; overall satisfaction; and costs of health care.

Orthognathic treatment is the process involving orthodontic treatment and surgery which is used to treat those patients who present with severe dentofacial deformities outside the scope of orthodontic treatment alone. Orthognathic surgery began early in the twentieth century but it was the introduction of the sagittal split technique which marked the introduction of "modern orthognathic treatment". Not only did the technique avoid an extra-oral approach but it also allowed either lengthening or shortening of the mandible. This was followed by the introduction of the Le Fort I downfracture technique and lead to the feasibility of correcting the majority of problems, even anterior open bites which had so far been impossible to treat to a stable position (Epker and Wolford, 1975; Wolford and Epker, 1975). Progress has continued so that it is now possible to treat dentofacial deformities of almost any type.

As both surgical and orthodontic techniques have progressed and more patients are undergoing this type of treatment, so it has become necessary for clinicians to analyse their outcomes critically. This may be done by a variety of techniques including:

- Assessment of surgical outcome in terms of aesthetics, function and stability.
- Assessment of patient satisfaction following treatment (Kiyak et al., 1982a, 1982b, 1984, 1986; Flanary et al., 1990; Finlay et al., 1995).
• Assessment of certain psychological aspects, for example, body image and self esteem (Kiyak et al., 1986; Lovius et al., 1990).

• Assessment of quality of life both pre- and post-treatment.

Health care providers in the UK are increasingly concerned with the measurement of the effectiveness of health care interventions. This is, in part, due to the competitive internal market which has been created and also in response to the necessity to meet customer needs and to provide “bench mark” standards of health care. As a result, patients' views are now seen as important measures of quality of care and health outcome (Hardy et al., 1996). Although clinical outcome measures are used routinely in dentistry, they may be of little importance to the patient. In contrast, preference based outcome measures take into account the patient’s lifestyle and well being (Matthews et al., 1999). Measuring patient based preferences stems from the economic theory that people are the best judges of their own welfare. A number of these so-called preference based measures have been used in medicine and have now been adapted for use in dental settings. These include clinical decision making which is a formalised method of transferring patient preferences to the clinician. It requires the use of two areas of information: risks and benefits of alternative treatment options and also individual “preference mapping systems”. The preferences of the patient can then play a part in choosing the most appropriate treatment option. Other preference based measures in the field of economic evaluation include techniques as diverse as utility measurement and contingent evaluation (Matthews et al., 1999).

Ronis et al. (1994) published a comprehensive review of patient-perceived outcomes of orthognathic surgery using an open-ended interview format. The authors proposed that some of these areas could be assessed by a single item (i.e. difficulty in opening the mouth) whereas others would require a more complex approach (i.e. self esteem, depression). The categories are shown in the following tables:
Table 1  Benefits of orthognathic treatment.

<table>
<thead>
<tr>
<th>Facial benefits</th>
<th>Dental benefits</th>
<th>Functional benefits</th>
<th>Psychological benefits</th>
<th>General benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better appearance</td>
<td>Straighter teeth</td>
<td>Easier eating</td>
<td>More confidence</td>
<td>Less pain</td>
</tr>
<tr>
<td>Better looking</td>
<td>Easier cleaning</td>
<td>Better digestion</td>
<td>More self-esteem</td>
<td>Eliminate jaw problems</td>
</tr>
<tr>
<td></td>
<td>Improved bite</td>
<td>Better breathing</td>
<td>Compliments</td>
<td>Full benefits of orthodontics</td>
</tr>
<tr>
<td></td>
<td>Overall better health</td>
<td>Improved speech</td>
<td>Treated better by others</td>
<td></td>
</tr>
</tbody>
</table>

Table 2  Risks of orthognathic treatment.

<table>
<thead>
<tr>
<th>Facial risks</th>
<th>Dental risks</th>
<th>Functional risks</th>
<th>Psychological risks</th>
<th>General risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swelling</td>
<td>Uncomfortable bite</td>
<td>Unable to eat</td>
<td>Worry about surgery, relapse, new</td>
<td>Pain</td>
</tr>
<tr>
<td>Bruising</td>
<td>Extractions</td>
<td>Weight loss</td>
<td>face, death</td>
<td>Wired jaws</td>
</tr>
<tr>
<td>Unrecognised face</td>
<td>Sensitive teeth</td>
<td>Reduced mouth opening</td>
<td>Fear of treatment</td>
<td>Numbness</td>
</tr>
<tr>
<td></td>
<td>Nonvital/broken teeth</td>
<td>Difficulty in talking, breathing</td>
<td>Depression, Stress</td>
<td>Infection</td>
</tr>
<tr>
<td></td>
<td>Teeth hard to clean</td>
<td>Joint problems</td>
<td></td>
<td>Reaction to anaesthetic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Feeling run down</td>
</tr>
</tbody>
</table>

Spilker (1996) divided outcomes of medical care into three broad areas which may also be used in the assessment of orthognathic treatment:

- Clinical: changes in signs, symptoms and directly measurable benefits.
- Economic: related to costs of care.
- Personal: the quality of life benefits the patients themselves perceive.
1.2 **CLINICAL OUTCOME MEASURES**

1.2.1 **Facial and dental aesthetics**

There have been a number of studies which have concentrated on changes in facial aesthetics following surgery. However, the changes in dental aesthetics have been looked at less frequently. The now widespread use of indices such as IOTN (Evans and Shaw, 1987; Brook and Shaw, 1989) and PAR (Richmond *et al.*, 1992) is likely to alter this in subsequent years.

The majority of studies which have looked at changes in facial aesthetics have used patients’ perceptions of their own aesthetic improvement and these were largely subjective opinions. Lack of standardised measuring techniques precludes other methods of assessment. Developments over the coming years are likely to concentrate on a standardised method of assessing changes in facial aesthetics such that a formalised study of outcomes may be undertaken rather than relying on subjective opinions of clinicians and patients.

1.2.2 **Function**

The number of studies looking at functional changes following surgery are far fewer than those studying aesthetic changes. This problem is further compounded by the very subjective nature of whether there is a functional problem and also whether this is altered following surgery.

Orthognathic surgery has the potential to alter occlusal force in at least two ways: mechanically by altering the jaw geometry and physiologically by changing the sensory and proprioceptive inputs or by altering other physiological variables. Proffit *et al.* (1989) studied the effect of orthognathic surgery on occlusal force during chewing, swallowing and maximum effort. Overall, larger changes were observed than could be accounted for solely by geometric alterations. Hunt and Cunningham (1997) also
reported marked alterations in occlusal force levels which continued to occur up to a year after surgery. They noted that advancement of the mandible resulted in weaker forces, whilst bimaxillary surgery for the treatment of a long face tended to bring previously weaker force levels to a more “normal” level.

1.2.3 Post-treatment stability (dental and skeletal)

Long term stability is an area of concern to both orthodontists and maxillofacial surgeons. A variable amount of dental and skeletal relapse is inevitably seen (Proffit et al., 1996).

Factors affecting dental stability may be divided broadly into: orthodontic mechanotherapy (for example, the proportion of pre- to post-surgical orthodontics); extraction pattern; occlusal interferences; periodontal factors and arch form. Likewise, factors affecting surgical stability may be divided into: type and magnitude of surgical movement; pre-operative facial morphology; type and duration of fixation and the surgical technique employed.

1.3 ECONOMIC OUTCOME MEASURES

Economic studies compare the costs of the intervention with the benefits derived from it and play an important part in the field of health policy and resource allocation. Studies may involve cost-utility, cost-benefit or cost-effectiveness techniques. However, to date, there has been very little research in this area in the field of orthodontics or maxillofacial surgery.
1.4 PERSONAL OUTCOME MEASURES

1.4.1 Psycho-social measures

The psycho-social outcomes of orthognathic treatment are of great importance to the patient and their friends/family. The impact of severe dentofacial deformity is still underestimated by the general public and often by clinicians. The following are extracts from an essay written by a patient who had undergone orthognathic treatment and describe some of the problems experienced during her childhood (Rachel, 1999):

"Now I was turning in on myself, my confidence shattered. I would cover my mouth when eating and laughing and shy away from contact with other children...."

"The thought of refusal to treat me...panic set in. Would my orthodontist reject me, as well as my schoolmates?"

"......a fresh round of name calling started and the school bullies joined in......"

Orthognathic treatment is now a very common procedure. However, the psychological implications of facial deformity and subsequent treatment are less clearly understood. Surgery produces sudden and sometimes dramatic changes which place immediate demands on the patient’s adaptation skills and they must learn to rapidly integrate these new facial features into their own self-concept. Patients undergoing surgery must be thoroughly prepared for the changes which will occur and if there is any doubt about the patient’s ability to cope they should be referred for a psychological assessment. Individuals who want “secondary gain” from treatment (for example, a better job or improved relationships) or who are seeking surgery due to external pressures (to please a parent or spouse) are particularly vulnerable and may experience problems in the post-operative period (Edgerton and Knorr, 1971; Peterson and Topazian, 1976). Clinicians must evaluate the patient’s psychological status carefully in order to be assured of the best possible outcome.
1.4.2 Patient satisfaction

Patient satisfaction is an important outcome and is now recognised by health care providers as a legitimate measure of health care quality. Pascoe (1983) defined patient satisfaction as “reactions to salient aspects of the context, process and results of their experience”. There are a number of reasons for measuring patient satisfaction (Hardy et al., 1996):

- Satisfaction is known to be associated with better health outcomes and patients are more likely to follow advice and treatment instructions.
- Customers (the patients) cannot express dissatisfaction by taking their custom elsewhere as with other consumer related issues. Therefore, it is important to establish how satisfied they are with health care.
- Recent years have seen greater competitive tendering within and between hospitals. This strategy helps to increase customer choice but unless the customers’ views are known the services are not truly consumer-based.
- Advances within medicine have led to greater clinician specialisation; less time spent in hospital following procedures etc. This may affect levels of patient satisfaction and the relevance of these aspects must be established.

Patient satisfaction has become a favoured parameter of post-treatment evaluation. An overall review of the literature suggests that the level of satisfaction following orthognathic surgery is between 92-100% (Flanary et al., 1985). However, if satisfaction is defined as the “willingness to re-elect surgery”, the level of satisfaction is reduced to between 84-92%.

A recent paper (Toivanen et al., 1999) used the familiar SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis to measure quality of public oral health services in Finland. Patients were asked what they expected from the service before treatment and afterwards were asked if their expectations had been fulfilled. Data were analysed in the conventional manner and also using the SWOT style analysis. The authors concluded that the SWOT analysis provided a structured interpretation of the results and may prove useful in assessing quality of services.
1.4.3 Quality of life

Quality of life measures have become a standard method of assessing results of health care interventions and of prioritising funding (Fitzpatrick et al., 1992). The term "quality of life" entered the literature between World War II and Lyndon Johnson's Great Society Programme, the concept being that a "good life" required more than just material affluence and possessions (Ferrans and Powers, 1985). The theoretical basis of many of the quality of life measures is that life gains its quality from the ability and capacity of the individual to satisfy their own basic needs (Hunt and McKenna, 1992). Cardiology, oncology and rheumatology were amongst the first specialities to undertake major quality of life research, probably due to the great impact of these conditions.

Quality of life is a difficult concept because different people place values on different things. There are a number of dimensions or domains which have been used in the study of quality of life and these include: the individual's opinion of their quality of life; socio-economic status; physical health; perceived stress; friendships and family; achievements of life goals; and satisfaction with self (Ferrans and Powers, 1985). Fitzpatrick et al. (1992) also divided the quality of life dimensions into several categories: physical, emotional and social function; role performance; pain and other symptoms (i.e. disease specific symptoms).

There are two basic methods of instrument designed to measure quality of life: condition-specific measure and generic measures (Guyatt et al., 1989; Garratt et al., 1996a). Both approaches have advantages and disadvantages and there are advantages in many cases to using the two in combination (Patrick and Deyo, 1989).
1.5 PSYCHOMETRIC PROPERTIES OF INSTRUMENTS

These issues apply to all three of the following chapters and are introduced here to avoid replication.

1.5.1 Validity

Validity may be defined as whether an instrument measures what it purports to measure. Streiner and Norman (1995) defined validation as “the process whereby we determine the degree of confidence we can place on inferences we make about people based on their scores from that scale”. Health services research is over-run with scales and measurements which have never been validated and it is this flaw which leads to the vast array of non-comparable results.

Until the 1970s most textbooks discussed the three “Cs” of validity: content validity; criterion validity and construct validity. Recent years have seen the advent of numerous subdivisions although the three basic terms still exist (St Leger et al., 1991; Streiner and Norman, 1995).

Content validity

This is the extent to which the instrument measures the domains under investigation. The higher the content validity of an instrument, the broader are the conclusions which can be reached. Content validity is generally claimed on the grounds that a number of representative judges were used to generate and select items and the resulting measure contains all the commonly mentioned items. It may also be demonstrated by reference to the literature. Content validity is sometimes used along with the term “face validity”. The latter term has less meaning and essentially means that it appears sensible to an intelligent audience.
**Criterion validity**

This is the extent with which the instrument produces measurements consistent with another independent instrument, ideally a “gold standard”. In using this type of validity it must always be questioned whether the existing instrument is valid. Often a new measure is being developed because existing measures are not acceptable, in which case this would not be an appropriate method. If a gold standard already exists, the only justification for developing a new measure would be that it is more practical, more economical or gives results sooner than the gold standard. Criterion validity is usually divided into two types:

- **Concurrent validity.** The criterion and the new measurement both refer to the same point in time. For example, the new measure and the gold standard would be given at the same time or within a short time of each other. This is the method which would be used if trying to replace an existing instrument with a simpler or cheaper one.

- **Predictive validity.** The new measurement predicts the criterion and, therefore, the criterion is not available until some point in the future. This is a very powerful indicator of overall validity and is useful in area such as diagnostic tests where there may be a time lapse waiting for results.

**Construct validity**

If there is no “gold standard”, construct validity may be used. This is the consistency of the measurement with theoretical concepts concerning the system under study. If a measurable attribute of a system should change with time, then the actual measurement should also change with time. This method of validity is often used when attempting to measure a hypothetical construct such as anxiety and, therefore, plays a major role in psychological research. Construct validity may involve comparisons with related measures and/or predictions about the distribution of scores in different groups.
1.5.2 Reliability

Reliability may be defined as "the process of determining the amount of random and systematic error associated with any instrument" (Streiner and Norman, 1995) or "the degree to which the measurement scores can be replicated" (McDowell and Newell, 1987). McDowell and Newell (1987) proposed that reliability is concerned only with random errors and that systematic errors (or bias) are assessed through validity testing. Poor wording can be a major source of error and should be checked thoroughly and piloted prior to using the questions in a definitive study.

Reliability may be assessed by the following techniques (McDowell and Newell, 1987; Streiner and Norman, 1995):

**Internal consistency**

Internal consistency is determined by statistical tests carried out on a single data set. This may involve a split-half method where the results of the two subscales can be compared. Other tests such as Cronbach's Alpha are based on the correlation between items included in the instrument.

**Inter-rater reliability**

This determines whether different raters, using the same method to assess the same respondent, obtain the same result. It is ideally measured by different raters interviewing the same respondents a short time apart. Typical values for inter-rater reliability fall into the 0.65-0.95 range. Values above 0.85 are generally considered satisfactory.

**Test-retest reliability**

This involves the patient completing an instrument twice, a short time apart. It must be assumed that the characteristic under test has not changed during the test period. It is, therefore, important to repeat the test long enough after the first so that the responses cannot be remembered but not so long after that there may have been any changes
which could affect the results. This method is commonly used in assessing the reliability of instruments used in psychology testing.

Assessment of inter-rater and test-retest reliability

There has been considerable debate in the literature regarding the most appropriate choice of reliability coefficient (Streiner and Norman, 1995). Many studies have used Pearson correlations, however, it must be recognised that a value of 1.0 can be obtained even if the intercept is not zero. In practice, however, the predominant source of error is random variation and under these circumstances, the Pearson correlation will usually provide an acceptable value. Streiner and Norman (1995) concluded that the Pearson correlation is theoretically incorrect but fairly close to the Intraclass Correlation Coefficient (ICC) which will yield a value of 1.0 only if the intercept is zero. Calculation of an Intraclass Correlation Coefficient involves looking at the sources of variation separately.

The difficult issues involved in reliability testing were also recognised by Bland and Altman (1986) who proposed a method whereby the differences between the two measurements and the standard deviation (SD) of the differences are calculated. A coefficient of reliability/repeatability can then be calculated using the equation below. Streiner and Norman (1995), however, criticised this method in that the coefficient is independent of the true variation in the sample of observations.

\[
BSI \text{ Coefficient of reliability} = 1.96 \times SD \text{ of differences}
\]

The methods which are quoted most frequently in the literature remain the Pearson correlation and the ICC and these are used in future chapters. The Bland and Altman method is quoted in addition for Chapter 2 as it allowed comparison with data in previous research.
1.5.3 Responsiveness

Responsiveness reflects whether changes in the relevant attribute are reflected by changes in the instrument/questionnaire score. An instrument must be responsive if it is to detect small but clinically significant changes in health status and the magnitude of changes on the scale should also reflect the magnitude of clinical change.
2.0 SUMMARY

In the past, performance measures in the United Kingdom have relied on areas which can be readily quantified, such as the number of finished consultant episodes. This method tended to support those who were treating more cases rather than those who were treating them to a higher standard. More recently, measures have been proposed which reflect changing concepts of health in a framework that has meaning to patients.

Patient based outcome measures have been popular in medicine for some years now but dentistry has been slower in adopting the technique (Corson et al., 1999). Indicators such as DMFS (decayed, missing and filled surfaces) still remain popular methods of assessing outcome, the major problem being that these do not take into account the patient’s own views of treatment. Outcome assessment can never be complete until patient perceptions are included.

Table 3 Uses of patient rated oral health indicators (Corson et al., 1999).

<table>
<thead>
<tr>
<th>USES</th>
<th>DENTISTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>• To judge whether the treatment provided has made any difference to individual patients.</td>
<td></td>
</tr>
<tr>
<td>• Clinical audit.</td>
<td></td>
</tr>
<tr>
<td>• To ascertain which patients are likely to benefit from treatment.</td>
<td></td>
</tr>
<tr>
<td>• Advocacy for health services in conjunction with clinical indices.</td>
<td></td>
</tr>
<tr>
<td>ADMINISTRATORS</td>
<td>• Monitoring clinical effectiveness.</td>
</tr>
<tr>
<td></td>
<td>• Targeting resources and service planning</td>
</tr>
<tr>
<td></td>
<td>• Cost benefit analysis.</td>
</tr>
<tr>
<td>RESEARCHERS</td>
<td>• Outcome measures in determining effectiveness of new treatments.</td>
</tr>
<tr>
<td></td>
<td>• Determining population subgroups with a shortfall in oral health.</td>
</tr>
</tbody>
</table>
CHAPTER 2

ECONOMIC EVALUATION OF ORTHOGNATHIC TREATMENT

“We never will have all we need. Expectation will always exceed capacity ...... This service must always be changing, growing and improving, it must always appear inadequate.”

(Aneuryn Bevin, 1948)
1.0 REVIEW OF THE LITERATURE

1.1 INTRODUCTION

Evaluation of health care programmes may be subdivided into evaluation of efficacy, effectiveness, efficiency and availability. The evaluation of efficiency is more commonly known as economic evaluation. Economic evaluation may be defined as "the comparative analysis of alternative courses of action in terms of both their costs and consequences" (Drummond et al., 1987). It is now a widely accepted tool for the appraisal of health care and this is reflected in the increasing number of research papers in this area in the medical literature. Economic evaluation sets out to answer two main questions. Firstly, is this health procedure worth doing compared with other things we could do with the same resources? Secondly, are we satisfied that the health care resources should be spent in this way rather than in any other way?

Economic evaluation is dependent on the quality of underlying medical evidence and, because of this, clinical trials are increasingly viewed as a natural vehicle for economic analysis (Drummond and Davies, 1991) although some have argued against this on the grounds that care in clinical trials is so different to normal practice that the data cannot be extrapolated (Evans and Robinson, 1980).

There are several issues which distinguish health care from other commodities (McGuire et al., 1988). Economics is concerned with choice and the study of economics assumes that the choices most consistent with maximising utility are those made under conditions of full information. This is one of the first issues which distinguishes health care, an individual suffering from an illness will often be unwilling or unable to collect the information to make this choice. Differences also exist when it is realised that health per se cannot be traded (although health care can), therefore "markets" in health do not exist. In addition, health care is consumed on the assumption that it will result in benefits in health status and individuals are not willingly engaged in health care consumption under normal circumstances.
CHAPTER 2 REVIEW OF THE LITERATURE

The relationship between health care and health status must be understood and accepted as a fundamental starting point for the study of economic analysis. The economist’s prime concern is health care not health *per se*.

1.2 ALLOCATION OF RESOURCES

As early as the 17\textsuperscript{th} Century, the British physician Richard Petty advocated greater social investment in medicine. This was based on his belief that the value of a saved human life far exceeded the cost (Torrance, 1982). Since the National Health Service was introduced in the United Kingdom, there has always been the basic problem that resources are limited and all the demands on the available resources cannot be met. Health care resources are limited by the total amount of resources available as well as through competition with other resources, such as housing and education. This raises the question of how to decide where the money should be allocated most appropriately. The establishment of a benchmark for an efficient level of health care provision is still to be found and it must always be questioned whether the allocation of health care resources is efficient and equitable.

Allocation of health care resources is a contentious issue. Health care and government agencies must decide how to allocate their resources for a wide range of very different interventions. This involves making difficult value judgments regarding the importance of certain health states. A number of arguments have been proposed in terms of “need” for and/or “right” to health care but problems arise when the amount allocated to a certain area or authority is insufficient to meet these needs. Allocation of resources is generally on two levels: planning and clinical (Carr-Hill, 1991). For planning decisions, this involves deciding whether or not facilities should be provided at all and, if so, where they should be located. Clinical decisions are made by practitioners on behalf of individual patients or groups of patients.
Goold (1996) also proposed two levels of organisation in order to ensure fair allocation of resources. She suggested that the first, based on the existing government structure, should be responsible for making broad decisions, such as the amount of money allocated to health care. The second level, made up of community organisations with a membership which represents a wide range of health related interests, should be involved in the making of policy decisions and the development of guidelines. Although this was a discussion aimed at improving allocation of resources in the United States, the main principles also hold true for the United Kingdom.

It is difficult to determine who should be responsible for the “rationing” of health care. Rationing of health care does not involve just medical decisions, there are also a number of moral issues which need to be considered. Some procedure therefore has to be established to allow the most appropriate allocation. This was the basis for the introduction of cost-utility analysis which assigns a ratio of cost to benefit and promotes efficient use of resources in a manner which is considered consistent with justice. The production of “league tables” in an attempt to establish which services give better value for money must, however, be treated with some caution (Gerard and Mooney, 1993).

Goold (1996) explored three ways of making decisions regarding the allocation of health care funds:

- Cost-utility analysis. This will be looked at in more detail later in this section. It is one of the most widely used outcome measures and is based on the principle of quality adjusted life years (QALYs).

- Informed democratic decision making (using the existing structure of government or a grassroots approach where citizens are involved in decision making). Although this may seem to be a fair approach, there are always certain subgroups which are liable to be under represented.

\[QALYs = \text{change in utility value due to intervention} \times \text{future life expectancy}\]
• "Veil of ignorance" or the theory of "prudent planners" (Rawls, 1971, 1993 cited in Cohen, 1996). This theory proposed that allocation of resources should start from the basic position of "behind a veil of ignorance" in which individuals do not know their future social or economic positions. It is claimed that from this point the most important principle of justice when making social choices is that absolute priority should be given to those who are worst off. In practice this would not always work, it may mean that vast amounts of resources are going to those who would benefit little (Cohen, 1996). Daniels' concept (1985, cited in Goold, 1996) is similar in that he proposed that prudent planners would restrict certain types of services which would benefit few people. These theories are unlikely to be useful on an everyday basis although they may be helpful in developing guidelines.

Autonomy has become important in the last two decades with patients and patient organisations contributing to the general undermining of the concept that clinicians have absolute authority (Jensen and Mooney, 1990). Autonomy may be attractive to doctors in the current climate where they have to ration health care. It has been proposed that, faced with increased demands but little increase in resources, the National Health Service has several options (Hine, 1999): to become more efficient so that more individuals can be treated with the same resources; to extend means testing so that some people may be excluded from certain services due to their wealth; and to increase rationing or to provide a smaller range of services. The way forward remains unclear.

1.3 COSTING HEALTH CARE

Health care is divided into costs and benefits. The benefits are usually health improvements which can be measured in a number of ways including:
• Health effects, for example, cases found, cases prevented, lives saved.
• Economic benefits, this can be measured in direct benefits (savings in health care costs because the programme makes the person healthier), indirect benefits
(individuals are able to return to work) and intangible benefits (monetary value of the reduction in pain and suffering).

- Value of the health improvement itself to the patient, family and society, regardless of the economic consequences.

Resources consumed can also be divided into direct and indirect costs as well as intangible costs (such as pain and suffering of the patient and their family). Costs in health care include costs borne by the NHS (staff, hotel services, drugs), costs borne by the patient and family (for example, travel) and costs to the rest of society (for example, health education). Robinson (1993b) classified costs as direct (staff wages), indirect (for example, loss of income due to illness) and capital (investments in buildings).

The real cost of any health care intervention is the loss of health outcomes from other programmes which have been forfeited by putting the resources in question into the first programme, this is known as the “opportunity cost”. Opportunity costs rest on the two principles of scarcity and choice. Scarcity means that societies do not have enough resources to meet all their citizens’ desires. As a result of scarcity, choices have to be made as to which activities a society should undertake and which should not be undertaken. Opportunity cost is of major importance to the economist and the aim of economic evaluation of health care services is to ensure that the benefits of the programmes implemented are greater than the opportunity costs of such programmes. Appropriate answers to two questions are also required (St Leger et al., 1991):

- What other technology could the money have been spent on, and what could that technology have earned?
- How else could the money have been invested and what would that have earned?
The principles of identifying, measuring and valuing costs are essentially the same regardless of what type of economic analysis is undertaken and a number of specific areas need to be considered (Drummond et al., 1987; Donaldson, 1990a):

- Identifying opportunity costs.

Items to be identified for inclusion in the cost side of an economic analysis are any resources which have an opportunity cost as a result of being used in the health care programmes under consideration.

- Which costs should be included?

Health care resource costs can generally be classified under four main headings: staffing; consumables; overheads (for example, administration or housekeeping costs) and capital costs (land, buildings and major items of equipment). Indirect costs are secondary costs related to paid and unpaid productive activities. Productive activities are those arising from participation in the labour force and from work within the home.

- How far and how wide should we go?

The question of whether future costs and benefits of programmes should be included is a difficult area. For example, in a breast cancer screening programme the disease may be detected at an earlier stage and lengthen life. However, people who survive may then develop other conditions in the future. Most economists would agree that the future costs and benefits should only be considered if they are closely related to the original condition, although this is an area which is open to debate.

- Counting costs in a base year.

Health care costs should be counted in a base year, this means they are adjusted to eliminate the effects of inflation. This has implications if a study is over a number of years. Whether all costs are adjusted to the first or the last year considered will not affect the final result.
• Discounting.

Not all costs and benefits occur at the same time. For example, costs of prevention are incurred early whilst costs of cure are incurred later. Most economists agree that costs (and benefits) occurring at different times should be weighted differently. Allowance needs to be made for the differential timing of costs and consequences, so-called “time preference” (Drummond et al., 1987). This relates to the fact that it is preferable to receive a benefit earlier or to incur a cost later, a cost arising in the future impinges on us less than an equivalent cost arising now. Future costs should therefore be discounted and given less weighting. Failure to do so can produce misleading results, therefore discounting future costs at a value around 6% is the norm (Torgerson and Raftery, 1999). Until recently it was common practice to discount both future costs and benefits, although discounting benefits has become more controversial (Torgerson and Raftery, 1999). It has been argued that health related benefits should not be discounted because health cannot be invested to produce future gain in the same way that money can. However, many economists still believe that benefits should be discounted, arguing that people value current health benefits more highly than future benefits. For this reason, the Department of Health (1995) suggested that benefits should be discounted at a lower rate, with 2% being commonly used.

• Patient based versus per diem costs.

Individual patient based costing is both time consuming and expensive. It is easier to calculate the per diem cost (cost per bed day) of providing a particular service and then multiply this by the length of stay. The main problem with this method is that the cost is not condition-specific but this can be overcome to some extent by calculating the cost per bed day for a certain ward where similar procedures are undertaken.

• Sensitivity analysis.

An important consideration in any economic analysis is whether a sensitivity analysis is required. Where it has not been possible to estimate resource use accurately, sensitivity analysis is useful as it involves testing the sensitivity of the results to variations in costs about which one is uncertain. Estimation of costs is difficult and
every analysis contains some degree of uncertainty, imprecision or methodological controversy. There are three main types of sensitivity analysis (Briggs, 1999). A one way analysis systematically examines the impact of each variable in the study varying it across a plausible range while holding all other values at their baseline value, it is likely that this method probably produces an underestimation of values. The second method is the extreme scenario which involves setting each scenario to take the most optimistic or most pessimistic value and may produce an overestimation. The third method is probabilistic sensitivity analysis and examines the effect on the results when the variables are allowed to vary simultaneously across a reasonable range.

Drummond et al. (1987) advised consideration of estimates made in an analysis which are either subject to debate because no estimate was available and an informed guess was made (i.e. costs of new techniques) or because of known imprecision in the estimation (i.e. per diem figures) or due to debate because of methodological controversy (i.e. choice of discount rate). Then upper and lower bounds on the possible range of estimates may be set and study results can be calculated based on combinations of the “best guess”, “most conservative estimate” and “least conservative estimate” of the variables concerned.

A further issue in cost estimates is that distributions are often skewed, for example, there may be a long right hand tail due to high costs related to complications or re-operation. For cost data, the crucial information is the arithmetic mean which is of importance in guiding healthcare policy but additional data are often quoted, in particular 95% confidence intervals (CI). Confidence intervals which are derived by conventional methods of analyses will only be reliable if the skewness is not extreme and the sample size is relatively large. The alternative is to use the bootstrap method which makes no assumptions about the shape of the distribution (Thompson and Barber, 2000). Thompson and Barber (2000) concluded that although conventional analyses are appropriate in most cases, results for skewed data or small samples should be checked using the bootstrap technique, which is a computer-based method.
for assigning measures of accuracy to statistical estimates (Efron and Tibshirani, 1993).

1.4 METHODS OF ECONOMIC EVALUATION

Drummond et al. (1987), Donaldson (1990a) and Robinson (1993a-e) discussed four methods of economic evaluation currently in use: cost-minimisation; cost-effectiveness; cost-utility and cost-benefit analyses.

Cost-benefit and cost-utility analysis both address the issue of outcome valuation and therefore shed more light on whether certain treatments are worthwhile. In contrast, cost-minimisation and cost-effectiveness assume that the intervention is worthwhile. It is, however, important to realise that none of these analyses can be used to replace sensible judgments but may be used as an adjunct to decision making. Care is also called for when studying papers which claim to use a certain analysis. Zarnke et al. (1997) studied papers which claimed to use cost-benefit analysis. Of 95 papers which met the inclusion criteria, only 30 (32%) met the definition of a cost-benefit analysis and the majority of the remaining papers were cost comparisons. Interpreting cost comparisons as if they were true cost-benefit analyses makes communication between health care researchers and policy makers very difficult and may have detrimental consequences when resources are allocated.

Cost-minimisation analysis (CMA)

This form of analysis is used when the outcomes of two procedures being compared are proven to be the same (for example, day stay or overnight stay treatment for removal of impacted third molars) and the aim of such an analysis is usually to find the lowest cost programme. A number of authors include cost-minimisation as a form of cost-effectiveness analysis.
Cost-effectiveness analysis (CEA)
Cost-effectiveness was the most widely used method of economic analysis until the 1980s. It answers the question “Given that it has been decided that this type of health care will be provided, what is the best way of doing so?” This method is used when the programmes may have differential success in outcome, as well as differential costs. The outcomes vary but can be expressed in common natural units such as life years gained or blood pressure reduction and cost-effectiveness is normally expressed as cost per unit effect. It is a useful technique for comparing alternative programmes whose effects are measured in the same units but it cannot be used to assess an isolated single programme and it is not possible to compare interventions which have several types of clinical effects. It was this disadvantage which lead to the development of cost-utility analysis (CUA).

Cost-utility analysis (CUA)
CUA should be the method of choice when quality of life is either an important outcome or the important outcome. It is the ideal method when the intervention affects morbidity and mortality or when treatments have a wide range of different outcomes and a common unit is required. Cost-utility analysis is said to lie somewhere between cost-benefit and cost-effectiveness analysis. “Utility” is a term used by health economists to refer to the subjective level of well-being that people experience in different health states.

The concept of utility developed from the work of the 19th century economist and philosopher Jeremy Bentham who developed the “utilitarian”1 school of thought. CUA can be used on two levels: the first is to evaluate a particular procedure or intervention and the second is to combine individual results into a “league table” as discussed later. Utility based measures are usually expressed in terms of quality adjusted life years (QALYs) which are weighted utility values. Information from QALYs, along with costs, can be used to guide resource allocation.

---

1 utilitarianism - the doctrine that actions are right because they are useful; doctrine that greater happiness of the greatest number should be the guiding principle of conduct (Concise Oxford Dictionary, 1982)
Cost-benefit analysis (CBA)

If the outcomes of two health programmes differ (for example, comparing hypertension screening with flu vaccination) then a common denominator must be established to allow comparisons of outcome. Cost-benefit aims to do this, usually in terms of money. CBA seeks to answer the questions “Is it worth achieving this goal?” and “How much more or less should be allocated to this type of health care?” It can look at one health care programme in isolation, although the alternative of doing nothing or continuing current practice is always implied (Donaldson, 1990a).

CBA has a solid foundation in the theory of welfare economics (Rittenhouse, 1995) and is one of the most comprehensive methods of economic evaluation which is available. It may take one of two approaches: the human capital approach or individuals’ observed/stated preferences, both of which will be discussed later.

Few studies have compared different methods of economic analysis for a health care intervention. However, Boyle et al. (1983) used cost-utility and cost-benefit analyses in their study of neonatal intensive care and Giesler et al. (1999) also used a combination of the two analyses in their study of patients with advanced prostatic cancer.
1.5 UTILITY ANALYSIS

Definition of utility:

"a cardinal measure of the strength of one's preference"

(Torrance, 1987)

1.5.1 History of the utility approach

The term "utility" has long been the subject of debate (Richardson, 1994). Initially the term was used to refer to a psychological concept of welfare and well-being and this view remained the popular one for many years. Even today, the literature often uses the term utility in the sense of a psychological and measurable quantity (Froberg and Kane, 1989a). The attraction of the utility approach as a measure of quality of life is probably derived from this interpretation. The second concept to emerge was based on the work of Fisher, Pareto and, more recently, Hicks and Allen (1934, cited in Richardson, 1994) who defined utility in terms of the order of preferences. This lead to the third concept in which utility was treated as an index of strength of preference rather than being simply an instrument to assist with the description and analysis of behaviour.

Throughout the 19th and 20th Centuries there has been a constant debate between advocates of cardinal (the belief that one can quantify the strength of preference i.e. I prefer A twice as much as B) and ordinal (the belief that one cannot go beyond saying I prefer A to B) utilities. Most of the literature around the 1930s discussed ordinal utility and it was not until the 1940s that cardinal utility was discussed again (Cohen, 1996). This renewed interest in cardinal utility was in response to the von Neumann and Morgenstern (vNM) theory of decision making under conditions of uncertainty (von Neumann and Morgenstern, 1953). The relationship between strength of preference and vNM utilities is also controversial and it is believed that vNM utilities reflect both strength of preference and attitude towards risk (Cohen, 1996). Although von Neumann and Morgenstern first introduced their theory in the 1940s, it was not
established in the field of medicine until the 1980s, having been first accepted into the field of business analysis (Jensen and Mooney, 1990).

1.5.2 Existing measurement techniques

The conventional utility scale is from 0 to 1 and is anchored by death at 0 and perfect health at 1. Utility of an outcome is different from the outcome itself. For example, Drummond et al. (1987) quoted the example of identical twins who both break an arm. One is a painter and one a translator, on a scale of 0 (dead) to 1 (perfect health) their rankings are likely to differ as each will assess the loss of arm movement differently. There are three commonly used methods for eliciting utility values (Drummond et al., 1987; Richardson, 1994):

Rating Scale (RS)

The score on a RS is the distance along a calibrated linear scale between two reference points and represents the value or worth which the subject gives to a particular health state. One of the main questions which the RS leaves unanswered concerns the functional relationship between the units of the scale and the utility or attribute in question.

The subject is asked to place the “best” and “worst” health states on the scale and then to place other health states in between. If death is judged to be the worst state and is placed at “0”, then the preference value for each health state is the number on the scale. However, if death is not judged to be the worst health state (constant pain and loss of physical functions may be judged to be worse) and is instead placed at a point “d”, then the preference values for the other health states can be calculated using the following equation (Drummond et al., 1987; McGuire et al., 1991):

\[(x - d) / (1 - d) \quad \text{or} \quad (x - d) / (100 - d)\]

(depending on whether a 0-1 or a 0-100 scale is used)
One of the main advantages of the rating scale is that the method is understood easily and various health states can be assessed simultaneously. However, there is no theoretical foundation for the technique. A danger in rating scale exercises is that context effects may influence the results. For example, valuations may be affected by other health states that are included. Bleichrodt and Johannesson (1997) found that the rating scale value for a health state did depend on the numbers of health states preferred and less preferred to it. They also showed that the severity of the health state had no impact as long as the number of preferred and less preferred health states remained the same. The authors proposed that the rating scale should only be used when measuring health states in isolation.

In 1996, Gudex et al. used the rating scale method in a large scale general population study with 3,395 respondents. A total of 45 health states were defined in terms of the EuroQol descriptive system (EuroQol Group, 1990) and each respondent valued a subset of these states using the RS method. The study suggested that members of the general public could distinguish between states of health that involved different degrees of severity. However, the values showed marked standard deviations which contradicted Torrance’s theory that values can be made more precise by increasing the sample size (Torrance, 1986). This study also found that social class and level of education had a significant effect on valuations. This was in contrast to a previous study which found no such relationships (Sackett and Torrance, 1978). If social class has an effect this may have important implications in the use of health state valuations.
The von Neumann and Morgenstern Standard Gamble (SG)

"...the theory tells us how individuals should make decisions under conditions of uncertainty based on a set of behavioural axioms and, in the process, how to measure individual preferences."

(Gafni, 1994)

The standard gamble method is said to be valid because it is based on the axioms of vNM utility theory for decisions under uncertainty. The outcome is actually a probability, $p$. The subject is offered Alternative 1 as a certain outcome versus Alternative 2, a lottery in which the probability ($p$) is varied until the subject is indifferent between the two alternatives.

Figure 1  The standard gamble technique.

This method has been accepted as the gold standard for many years although Richardson (1994) argued that the principles from which the von Neumann and Morgenstern utility was derived are empirically flawed. The author proposed that the SG should no longer be regarded as the gold standard but rather as one of several scaling techniques to be evaluated on an equal basis with the other techniques. A
number of other arguments have been proposed against the SG technique. It is said to be a technique which may be difficult for respondents to understand and is also sensitive to the framing of the questions (Kahneman and Tversky, 1979). It has also been said that strong interval criterion would only be met if the von Neumann and Morgenstern principles were correct. In addition, the risk in SG is an unrealistic situation (although at least there is some element of risk).

Despite these arguments, the SG is still widely used in utility analysis. SG based QALYs are accepted units because they do reflect choice; the probabilities fulfill interval criteria; the units are readily understood and the SG reflects risk, something which is involved in any medical intervention.

**Time trade-off (TTO)**

Torrance *et al.* (1972) originally developed the TTO because some respondents found SG probabilities difficult to understand. The TTO asks the subject to imagine that they personally experience the health state and the time in that health state, $x$, is varied until the respondent is indifferent between the two alternatives (state $i$ and perfect health). The diagram below shows the application of TTO for a chronic health state considered better than death. Stiggelbout *et al.* (1995) proposed that the only time period which should be used in the TTO is the subject's life expectancy.

**Figure 2** The time trade-off technique.
One of the main problems associated with these three methods is that people are valuing their own health and, from this, valid conclusions cannot be drawn about how society wishes to balance health benefits for different groups. The person trade-off (PTO) was suggested as a method of avoiding these problems and gives the choice “If there are x people in adverse situation A and y people in adverse situation B, and you can cure one group, which would you choose?” Pinto Prades (1997) compared the SG, RS and PTO and proposed that the PTO performed slightly better than the SG. However, as yet, it has not become an accepted method.

Other methods include:

- Magnitude Estimation (ME): This technique is less commonly used (Richardson, 1994) and asks subjects “How many times worse is one health state than another reference state?” There is no universally accepted scale for health states and this makes it difficult to place meaning on the responses.

- Healthy-years equivalents (HYE): This method was proposed by Mehrez and Gafni (1989) and involves converting the time in ill health to the time in full health. The authors proposed that to measure HYEs required a two-stage procedure: the standard gamble to measure the utility of the health state and the utility is then equalised to the number of years in full health. However, Johannesson (1994) suggested that HYEs could be measured with the same degree of accuracy using a one-stage TTO method. At the present time, there has been relatively little research into the HYE and the reliability and validity have not been investigated thoroughly. In addition, the interviews tend to be more complex and time consuming which may influence how many subjects are willing to take part in a study (Mehrez and Gafni, 1991).
1.5.3 Comparison of methods

A common criticism of the rating scale and the standard gamble is that they require two separate functional relationships: the first between the health state and the scale and the second between the scale and the unit of output. The appeal of the TTO is that these two relationships are combined into one and the health state to be evaluated is contrasted directly with a reference state which acts as a measurement unit (for example, the TTO measures the number of healthy years that are considered equivalent to a given period in a particular health state).

All three methods have proved practical on most populations but in view of the fact that the RS may be measuring aspects of health status rather than truly valuing health states, the SG and TTO are recommended (Brazier et al., 1999). Badia et al. (1999) noted that all available methods suffered, to some extent, from problems of inconsistent responses. These inconsistencies may be due to respondent characteristics or may be due to some aspect of the measurement procedure. Giesler et al. (1999) found that the TTO gave the greatest number of inconsistencies and, in contrast to the previous study, concluded that all methods performed at a sub-optimal level.

Torrance (1976) assessed feasibility, reliability, validity and comparability of all three methods. Subjects ranked the TTO as the easiest and, surprisingly, the RS last. However, the SG was used only with a selected group of respondents who were considered “highly educated” and were, therefore, more likely to understand the concepts involved. The interviewers had no preference between the methods but acknowledged that the SG and TTO tended to be more expensive due to being used most frequently in an interview situation. Reliability was tested in two ways. Internal reliability was assessed using a different description of the same condition later in the interview and replicating measurements. Internal reliability of the three methods was generally good with values between 0.86 and 0.94 for the RS (Torrance et al., 1982), 0.77-0.96 for the SG (Torrance, 1976) and 0.77-0.88 for the TTO (Torrance, 1976; Torrance et al., 1982). A test-retest was also undertaken one year later. The test-retest gave lower coefficients of reliability (RS 0.49; SG 0.53; TTO 0.62) but there
was no significant difference between the three methods. However, it must be borne in mind that the period between the two tests was longer than normal. Using the SG as the gold standard for criterion validity, the TTO gave a coefficient of validity of 0.65 (0.84 after correction for internal unreliability) but the rating scale was unsatisfactory with a coefficient of 0.36. All three methods were tested for equivalence and the RS was significantly different to both the SG and TTO but comparison of the SG and TTO did not show significant differences.

Validity of utility values was also examined in a later paper by Torrance (1987). If the standard gamble is accepted as the gold standard then other values can be compared against the utility values derived from this method. It has been suggested that the rating scale values should be related to the TTO and SG not by a linear function but by means of a power curve. A second method of ascertaining validity has also been proposed, the theory is that health state utilities are said to measure overall quality of life, therefore, should be strongly correlated with other trusted HRQL measures (Torrance, 1987).

Reliability can be expressed in terms of individual measurements using the standard deviation of the differences in measurements. Standard deviations of the measurement error in the region of 0.09 to 0.15 have been quoted for the rating scale (Torrance et al., 1982) and values around 0.13 for the SG and TTO (Torrance, 1976). Sackett and Torrance (1978) also found large individual variations. There is, therefore, little doubt that single individual utility measurements are relatively imprecise and it is interesting to note that, with a few exceptions, these differences cannot normally be explained by patient characteristics or demographic variables. However, it is encouraging that mean values for groups of individuals are relatively stable (Sackett and Torrance, 1978; Boyd et al., 1982). Measuring utility values by several different techniques and on several different occasions should minimise error.

Utility values for the same health state have also been found to vary across different groups. Torrance (1976, 1986) found that subjects who were more knowledgeable
about the health state appeared to show lower standard deviations. This was indicated in their study of home dialysis by standard deviations of 0.3 and 0.2 for general public and patient groups respectively (Torrance, 1976).

It is apparent that different utility values have been produced by the different methods and values also differ from study to study. McGuire et al. (1991) proposed that this may be due to a number of reasons including that health states may be described differently in different studies. It must be decided whether to use specific descriptions or generic descriptions. Values may also vary depending on who is providing the assessment (for example, current patient, past patient, clinician or general public) and on whether prognostic information is given. One of the important aspects when considering errors in utility measurements is that systematic bias may be introduced, depending on how the questions are phrased (i.e. whether the subject perceives the situation as a loss or a gain, or whether death is implicitly mentioned).

1.5.4 Sources of utility values

Health state utilities are measured either by patients who are in that health state or by raters who may or may not have experience of the health state (Torrance, 1987). Using patients avoids the problems of describing a particular health state and has the advantage of combining two research tasks in one but may give biased values. In addition, there has been resistance from clinicians who worry about the additional stress placed on patients in undertaking a direct utility measurement approach (Brazier et al., 1999). Torrance (1986) proposed that the choice of respondents can be determined in part from the purpose and viewpoint of the study. For example, if an assessment of breast cancer screening alone was being undertaken, then those at risk may be the most appropriate respondents. However, if breast cancer screening was being compared with renal dialysis, community values may be more appropriate.
Participation rates in utility studies are lowest for the general public (70-84%) and highest for those with a special interest, for example, patients and clinicians (83-100%). Completion rates have been found to be high for both groups and 96% or greater of those who start an interview complete it. A number of studies have investigated whether there are variations depending on demographic variables such as age, gender and ethnic group. Fortunately, it has been a general finding that utility values are relatively stable across different groups and Torrance (1982, 1987) found that group mean utility values were remarkably stable regardless of the make-up of the group.

Sackett and Torrance (1978) undertook one of the early health care utility studies in an attempt to derive utility values for different health states as perceived by the general public. At this stage, health state utilities had been determined by governments, health professionals and voluntary health organisations but not by the general public. Utility values were determined using the TTO technique for a range of different health states. The relationship between health state utilities and age, gender and social class were also investigated. The authors found that six of the fifteen health state utilities were statistically significant with respect to age but none differed significantly with respect to gender and socio-economic status. This study concluded that the general public could generate usable health state utilities and this set a precedent for incorporating values of the general public into decision-making about health care programmes.

The following table illustrates health state utilities from a general population sample (Sackett and Torrance, 1978).
Utility values can be obtained from three sources (Torrance, 1986). The first, and simplest, method involves estimating values. This judgement can be made by the analyst or by a panel of experts. If this method is used then extensive sensitivity analyses should be undertaken. Secondly, there is also a growing list of health states which have utility values quoted in the literature, although it is vital that all criteria are identical if utilities are used from the literature. The third method is direct measurement and this is obviously the most accurate method. It may be the subject’s own health state which is being measured or a description of a health state may be given and the subject is asked to estimate a utility value for it.

The required sample size for economic evaluation studies is frequently difficult to establish because the number of previous studies is limited. Whilst some studies have
used very large samples, others have produced useful data on much smaller samples. For example, Hornberger et al. (1992), in their dialysis study, calculated that an effect size of 20% between two methods of measurement would be detectable in a sample size of 58 patients. Hall et al. (1992) estimated that 60 patients and 44 members of the general public was an adequate sample size to estimate breast cancer utility values using the TTO. Torrance (1976) used sample sizes between 29 and 43 in early utility studies.

1.5.5 Description of health states

If raters, rather than patients, are assessing health states it is important that clear descriptions are provided. The health states should be described in functional or behavioural terms rather than clinical terms. Torrance (1987) described six areas which should be included: physical function; emotional function; sensory function; cognitive function; self-care and pain.

When describing a health state, it is important to specify whether the health state applies to the respondent or to someone else. The most common method is to instruct the respondent to consider the health state applies to them (Sackett and Torrance, 1978; Torrance, 1982). There is still insufficient evidence regarding the level of detail to be included in the health state description. Some investigators include only a few key words, others include lengthy descriptions and a few have made use of video or audio tapes. In general it is accepted that the health state description should contain a maximum of 6 or 7 key aspects. Whether the prognosis of the state should be specified is a controversial area. Some researchers claim that prognosis is not part of a health state description, however, Torrance (1982) recommended that the duration and prognosis of the health state should be specified.

A number of factors may influence utility values (Froberg and Kane, 1989b; Ferguson and Keown, 1995). These include: the subject’s own experience in the health state; the
actual or projected duration of the health state; the description of the health state; anchor points on the scale and framing of the health outcome as loss or gain (for example, an outcome may be described as a 30% chance of death or a 70% chance of recovery). There has been very little research into the validity of different forms of health descriptions and a diversity of formats have been used. Gerard et al. (1993) studied framing and labelling effects in their study of mammography screening. Subjects were randomly allocated to nine different presentations of two basic breast cancer descriptions and analysis of the results showed that there was no overall significant effect of framing or labelling variables. The use of the word “cancer” did influence subjects’ valuations negatively but overall there was no significant difference. It may be that the effects are so subtle that the method was not able to detect them or it may be an indicator of the reliability of health state descriptions.

1.5.6 Advantages and disadvantages of utility analysis

Utility analysis is useful in that it allows comparison of a wide range of different interventions including those which are life saving and those which are life enhancing (Gerard et al., 1993). The relationship between quality and quantity of life as described by Torrance and Feeny (1989) is shown on the following page.
Health care interventions aim to create some improvement in life expectancy and/or quality of life. QALYs create a way of measuring both of these and reflect that most people will sacrifice some quality of life for quantity and vice versa (Williams, 1995). Thus QALYs are a measure of whatever aspects of life the individual values. QALYs developed as a concept in the 1970s and it was in 1977 that Weinstein and Stason (cited in Torrance and Feeny, 1989) first published a paper on CEA, describing the QALY as the appropriate measure of effectiveness. Utilities are not the same as QALYs, but are used as a weighting factor to adjust the remaining life years for the quality of life which will be experienced (Bakker and van der Linden, 1995).

The advantages of the utility approach are that it is generalisable, comprehensive, can integrate morbidity and mortality effects, has the ability to represent multiple viewpoints by using different evaluators of health states, produces a single score and is consistent with economic evaluation. Much of the appeal of utility analysis is attributed to the fact that its unit of output is the QALY. Most people would agree that quality of life is an important consideration in the allocation of resources and the...
emphasis in this area has increased dramatically over the last few years. The concept of applying differential weightings to QALYs, for example, higher weightings for young people over older people, or more seriously ill over less seriously ill has also been introduced more recently (Williams, 1995; Nord et al., 1999).

However, the utility technique is not without problems and Mooney (1994) discussed some of these issues:

- QALYs do not allow for different health preferences and some individuals may value health more highly than others.
- Health states are valued equally, regardless of the individual involved. For example, a violinist is likely to value loss of an arm more highly than a singer.
- There is no allowance for “diminishing marginal utility of health”. As an individual receives more health gains, the benefit from each additional health gain is likely to diminish. In addition, gains in health are not affected by the health status before intervention.
- There is no independence of health states across different time periods. Therefore, extreme pain for two seconds is one-eighteen-hundredth of what it would be if the extreme pain lasted for one hour.

Richardson (1994) also highlighted some of the difficulties associated with this technique when he questioned whether “quality” always has the same meaning and whether different sampling techniques may be measuring different concepts of quality. He questioned whether the techniques used actually measured the concepts of utility as described in the literature and whether these concepts of utility formed a sound basis for decisions regarding allocation of resources.

The main disadvantages of this method are that utility measures vary considerably between individuals, the interviews are labour intensive and utility scores are not always readily interpreted. These issues mean that care must be taken in how utility values and QALYs are presented and further research into how these problems may be overcome is essential. Despite these disadvantages, CUA is a popular technique and is
likely to remain so until a more appropriate method becomes available. Although QALYs cannot be used to make judgements about the total health care for a community, it is a useful technique for allocating resources within a given budget (Mooney, 1994).

1.5.7 QALY league tables

QALY league tables were devised to have a role in priority setting of health care programmes which were competing for resources. Procedures are ranked in order of cost per QALY gained with the underlying rule being that programmes should be implemented on the rank order basis of ascending cost per QALY. These tables have caused a great deal of controversy amongst policy makers and clinicians alike and the suggestion by Mooney (1994) and Drummond et al. (1995) that the tables do have a use but should be used with great caution seems a sensible one. Williams (1995) stressed the point that we can “no longer cling to the notion that if only we had more resources devoted to health care, priority setting would not be required.” We are faced with a situation where priority setting is required and league tables may be used as a guide.

There are numerous problems associated with the use of QALY league tables. The Oregon Health Services Commission set out to create a prioritised list of health care services which would be covered under the Medicaid programme (Hadorn, 1991). Their draft priority list was subsequently heavily criticised in that surgery for ectopic pregnancy and appendicitis were rated below dental crowns and splints for temporomandibular jaw dysfunction, respectively. This led to the development of an alternative method of prioritisation which produced a more realistic final list. League tables also differ between different areas and are basically a reflection of what is happening in that specific area. Therefore, policy decisions cannot be made based on a league table from another geographical area without first establishing that the local situations are transferable (Mooney, 1994). At present, there has been little research
on the transferability of clinical and economic data from one area to another (Mason and Drummond, 1995). A further complicating factor is that QALY league tables still remain incomplete and many procedures have not been subjected to CUA.

McGuire et al. (1991) discussed some of the other particularly controversial areas including the issue that some would argue that gains in quality of life and gains in length of life should be represented in separate tables. It may also be more appropriate to have separate tables, for example, for elderly patients and for children. There is also the issue that tables may imply that those interventions with a very high cost per QALY would not get treatment at all. The following example of a QALY league table is adapted from work by Culyer (1991, cited in Petrou and Renton, 1993):

**Table 5  Example of a QALY league table.**

<table>
<thead>
<tr>
<th>INTERVENTION</th>
<th>COST PER QALY GAINED (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign intracranial tumours</td>
<td>240</td>
</tr>
<tr>
<td>Pacemaker implantation for heart block</td>
<td>700</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>750</td>
</tr>
<tr>
<td>CABG for severe angina (left main disease)</td>
<td>1,040</td>
</tr>
<tr>
<td>GP control of serum cholesterol</td>
<td>1,700</td>
</tr>
<tr>
<td>CABG for severe angina (two vessel disease)</td>
<td>2,280</td>
</tr>
<tr>
<td>Kidney transplant</td>
<td>3,000</td>
</tr>
<tr>
<td>Breast cancer screening</td>
<td>3,500</td>
</tr>
<tr>
<td>Heart transplantation</td>
<td>5,000</td>
</tr>
<tr>
<td>CABG for mild angina (two vessel disease)</td>
<td>12,600</td>
</tr>
<tr>
<td>Hospital haemodialysis</td>
<td>14,000</td>
</tr>
</tbody>
</table>

NB: 1983-1984 prices
1.5.8 Studies involving the utility approach to assess chronic health states

A number of studies have obtained utility values for chronic, non-life threatening conditions which may be comparable to orthognathic treatment. In their 1996 study, James et al. carried out a utility analysis of orthopaedic treatment in a hospital in the north west of England in an attempt to establish whether it was possible to prioritise elective care. The study developed utility values for a range of orthopaedic procedures using the Rosser classification (Gudex and Kind, 1988) and the EuroQol (EuroQol Group, 1990). The issue was raised as to whether large but expensive benefits, such as hip replacements, could be justifiably compared with small benefits at low cost, for example, carpal tunnel repair. The authors proposed that this could be overcome by ranking procedures as major, intermediate and minor.

Other diverse chronic conditions have also been valued. For example, Cheng and Niparko (1999) conducted a meta-analysis of cochlear implants in adults and found utility values for profoundly deaf adults of 0.54 before implants and 0.80 following surgery. The authors noted that the published studies had used the rating scale rather than the SG or TTO. Palmer et al. (1999) also looked at cost-utility of cochlear implants and concluded that “although the link between hearing impairment and reduced health is multifactorial, one hypothesis suggests that effective interaction across a diverse social network may carry favourable health effects”. Using the TTO technique, Sculpher (1998) found a mean utility value of 0.50 for patients suffering from menorrhagia, with values of 0.86 following hysterectomy. Wells and Sherbourne (1999) used the TTO and SG to assess utility values for a range of chronic medical conditions and found that depressed patients placed a lower utility on their current health than patients with common chronic medical conditions.

Mittmann et al. (1999) used the Health Utilities Index Mark III to determine utility values for a wide range of chronic conditions in a Canadian population. The mean values for the conditions are shown in the following table.
Table 6 Utility values for chronic health conditions.

<table>
<thead>
<tr>
<th>HEALTH CONDITION</th>
<th>MEAN UTILITY VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>No chronic condition</td>
<td>0.93</td>
</tr>
<tr>
<td>Acne</td>
<td>0.92</td>
</tr>
<tr>
<td>Alzheimer's disease</td>
<td>0.58</td>
</tr>
<tr>
<td>Arthritis/ rheumatism</td>
<td>0.78</td>
</tr>
<tr>
<td>Asthma</td>
<td>0.86</td>
</tr>
<tr>
<td>Back pain</td>
<td>0.81</td>
</tr>
<tr>
<td>Bronchitis/ emphysema</td>
<td>0.79</td>
</tr>
<tr>
<td>Cancer</td>
<td>0.82</td>
</tr>
<tr>
<td>Cataracts</td>
<td>0.77</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.79</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>0.78</td>
</tr>
<tr>
<td>Food allergy</td>
<td>0.85</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>0.79</td>
</tr>
<tr>
<td>Heart disease</td>
<td>0.77</td>
</tr>
<tr>
<td>High blood pressure</td>
<td>0.82</td>
</tr>
<tr>
<td>Migraine</td>
<td>0.83</td>
</tr>
<tr>
<td>Other allergy (not food)</td>
<td>0.88</td>
</tr>
<tr>
<td>Sinusitis</td>
<td>0.84</td>
</tr>
<tr>
<td>Stroke</td>
<td>0.68</td>
</tr>
<tr>
<td>Stomach/ intestinal ulcer</td>
<td>0.80</td>
</tr>
<tr>
<td>Urinary incontinence</td>
<td>0.70</td>
</tr>
</tbody>
</table>

1.6 THE COST-BENEFIT APPROACH

"...political economy has to take as the measure of utility of an object the maximum sacrifice which each consumer would be willing to make in order to acquire the object ....... the only real utility is that which people are willing to pay for."

Jules Dupuit (1844, cited in Donaldson, 1993)

1.6.1 Introduction

Weisbrod et al. (1980) proposed that “a cost-benefit analysis should be seen not as a mechanism for deciding mechanically on the allocation of funds and resources among programs but as a structure for weighing advantages and disadvantages (that is, for organising knowledge)".
Cost-effectiveness and cost-utility analyses are limited in that they offer no guidance to the decision maker about the value of the foregone alternatives when a programme is implemented. To demarcate a cost-effective from a cost-ineffective programme, the decision maker must decide the maximum they are willing to spend to achieve the specified outcome. Cost-benefit analysis assigns monetary values to costs and outcomes and the programme’s net benefit is easy to calculate, thus giving the decision maker a single measure of the effectiveness of the programme. Advocates of CBA claim that this helps to facilitate decisions (Zarnke et al., 1997). The aim of cost-benefit analysis is to establish whether a number of investment projects should be undertaken at all and, if available funds are limited, which projects should be selected (Gafni, 1991).

Cost-benefit is the application of the theory of resource allocation. It appeals to economists because it is based on normative welfare theory which states that individual consumer preferences regarding cost-outcome should be adhered to in resource allocation (O’Brien and Gafni, 1996). In contrast, the underlying theoretical basis of cost-effectiveness is not clear (O’Brien and Viramontes, 1994). Much of the history of welfare economics has been dominated by social welfare. By the 1930’s this had largely been rejected and policy changes were seen which were “Pareto improving”, that is changes would be implemented if one person was made better off without making anyone worse off (Mitchell and Carson, 1989). The theoretical basis of cost-benefit analysis is linked with the Pareto improvement or rather the “diluted” version, known as the potential Pareto improvement (or Pareto criterion). This may be defined as “an economic rearrangement in which the gains can be so distributed as to make everyone in the community better off”, that is, the gainers can more than compensate the losers (Mishan, 1988; Gafni, 1991; Donaldson, 1993).
A cost-benefit calculation per se shows that the total gains exceed the total losses. However, the calculation may be accompanied by observations and recommendations, for example, if a certain group would suffer considerable loss then the policy makers would have to question the wisdom of introducing such a scheme (Mishan, 1988).

There are two approaches to cost-benefit analysis:

- **Human capital approach:** This is based on the principle that humans can be expected to continue to work in future years so benefits of health care can be measured in terms of the future income which would otherwise be lost due to illness. This technique obviously has some restrictions particularly when elderly or unemployed patients are being considered (Donaldson, 1990a). In addition, a programme which treats minor ailments in high earners may appear to perform better than one which treats severe problems in low earners. There is some debate as to whether the human capital approach is consistent with the theoretical foundation of CBA from a welfare viewpoint (Zarnke et al., 1997).

- **Individuals' observed or stated preferences:** One well-known example of this is the willingness-to-pay approach, a technique which has been used more widely in environmental issues than in health care to date.

### 1.6.2 Willingness-to-pay/ Willingness-to-accept

There are a number of commodities for which no market value exists. The contingent valuation method assigns monetary values to such commodities and this has led to a renewed interest in the cost-benefit analysis (Kartman et al., 1996). Contingent valuation was first used in the field of environmental issues but has been introduced more recently to health care. Economists are used to measuring the value of goods which are bought and sold in the marketplace. However, it is more difficult to value so-called public goods such as clean air or clean water (Mitchell and Carson, 1989). In the United Kingdom, health care is also difficult to categorise. On one hand, it may be perceived as being free but on the other hand, many patients feel that they contribute to costs because they "pay their stamp".
Contingent valuation includes the willingness-to-pay (WTP) and the willingness-to-accept (WTA) approaches. Of these, the WTP method is the most commonly used. Donaldson, in his 1996 Doctoral Thesis, found only 28 WTP studies for health care in English language sources and in none of these studies were WTP values elicited for more than one type of health care intervention. Therefore, there is no suggestion of how a particular service performs with respect to other services (Donaldson, 1996, cited in Olsen and Donaldson, 1998). Willingness-to-pay for a service and willingness-to-accept compensation to forego the same health care intervention have been shown to be consistently different. WTP will always be limited due to constraints on an individual's budget but WTA compensation is higher because respondents believe that the compensation budget has fewer constraints (Mishan, 1988; Donaldson, 1993). Another theory to explain the disparity between WTP and WTA is that an individual who has access to a service is likely to “protest” at having to give it up and will claim a higher amount of compensation. Mitchell and Carson (1984, cited in Donaldson, 1993) proposed a property rights approach to overcome this situation. They proposed that health care is a non-transferable collective good which makes WTA an irrelevant concept and that the only alternative is to ask people their WTP.

There are two approaches involved in WTP studies (O'Brien and Viramontes, 1993; Chestnut et al., 1996):

- The indirect, or averting behaviour, method which infers WTP from real life decisions involving trade-offs between money and expected health outcomes (for example, increased wages accepted by workers in occupations with known health risks). This is a difficult method to apply because instances of trade-offs between money and health are rare.
- The direct method which involves asking people to estimate the maximum amount they would be willing-to-pay to obtain a service. This is the more commonly used method.
There has been growing interest in the WTP method as a result of some of the limitations of QALYs and it is now seen as an important and developing technique within the field of health care research (Brazier et al., 1999). QALYs may vary depending on the nature of the service and there are also non-health benefits from health care use, for example provision of information, which are not readily accounted for. Unlike QALYs, WTP does not place the same constraints on which attributes can be valued (Olsen and Donaldson, 1998).

The WTP technique asks patients what they would be willing-to-pay to obtain the benefits of a particular health care intervention. This is usually performed with postal surveys, questionnaires or interviews, although recent WTP recommendations in the field of environmental benefits state that interviews should be used in preference to other methods (Arrow et al., 1993). One area of difficulty is that those who are affluent will be able to afford more in their WTP whilst those who have less money will be more readily compensated. It therefore ignores the considerations of distributive justice (Donaldson, 1993).

Willingness-to-pay is a particularly useful technique in cases where consumers are not being asked about health gain but are being asked to choose between two different health care choices (Donaldson, 1990b; Donaldson et al., 1995a). The technique can be used to provide data regarding strength of preference of treatment options and also the value placed on individual treatment methods. This can then be used alongside other economic evaluation data and allows decision making to be based on accurate assessments of people's values rather than what the policy makers believe is correct.

Individuals cannot plan to use health care services, therefore it has been suggested that WTP questions should be phrased as the maximum amount they would be willing-to-pay as an insurance premium so that the service is available if needed. The respondent, therefore, has to be informed of the probability of ever having to use the service. Morrison and Gyldmark (1992) reviewed a number of WTP studies and found that none posed the question as an insurance question. Donaldson (1990b) came closest by
asking how much respondents would be prepared to pay in extra income tax. Studies have tended to ask those who are using or will use the service rather than posing the insurance question. This form of selective sampling provides information that could be used to allocate resources in a way that maximises the aggregate utility of current users. However, if social welfare is of interest, selective sampling defeats the purpose of using this method. Values are usually elicited from those who belong to the relevant population unless altruistic values are of interest. (Morrison and Gyldmark, 1992). Gafni (1991) concluded that WTP questions should be presented to all persons who are affected by the availability of the programme not just current patients. This means asking current users, those who may need the service in the future and also those who experience non-user or externality effects (for example, friends and families of users).

There are several possible methods for measuring WTP (Mitchell and Carson, 1989):

- Bidding game. This is the oldest and most widely used method. It has the disadvantage that the final amount may be influenced by the starting amount.
- Payment card method. This was developed as an alternative to the bidding game and avoids having to give a starting bid. A range of realistic values are presented, ranging from zero to a high figure.
- "Take it or leave it approach". A large number of prices are set and the respondent is asked if they would be prepared to pay one certain price. This has the disadvantage that there are those respondents who will automatically say yes (the so-called "yes bias").
- Open-ended method. The respondent is asked how much they would be prepared to pay without being given any guide. Many respondents find this a difficult concept, particularly in the United Kingdom where patients are not used to the costs involved in health care.
- Split sample. Different methods can be used in the same study if the sample size is of sufficient size.
1.6.3 Advantages and disadvantages of WTP

One of the most common arguments against QALYs is that they permit the valuation of health gains only and, whilst this is obviously a major part of health care, health care is multidimensional and aspects such as "process" of care are not taken into account. WTP offers a way of including all of these attributes. It also allows a method of assessing choices where health gain is not the major issue, for example, when information is being provided by screening services (Donaldson et al., 1995a). The concept of "process" of care is controversial, it is difficult to define and even more difficult to measure (Gerard and Mooney, 1993; Donaldson and Shackley, 1997). Although it is assumed that health care does not have positive values in use, this argument ignores the negative values, for example, laparoscopic surgery may be preferred to more invasive surgery and taking a tablet may be preferred to an injection (Donaldson and Shackley, 1997). It is this aspect which QALYs are unable to take account of but which the WTP method may incorporate.

There are, however, several controversial issues with regard to the method. Firstly, the association of willingness-to-pay with ability to pay. Donaldson et al. (1995a) investigated preferences regarding obstetric care in Scotland and found that the maximum WTP was not associated with indicators of ability to pay but that the results reflected the strength of the preference alone rather than being a reflection of the ability to pay. They concluded that the relationship between willingness and ability to pay need not count against the method.

Another problem area is that respondents may object to or misunderstand the concept. It is important that the respondents be told that the situation is hypothetical. However, that in itself may lead to an overestimation of the preferred amount (Donaldson et al., 1995a). Many studies have found that respondents object to the concept by registering a "protest" and feel they should not be made to consider paying even in a hypothetical situation. A third controversy is whether to use an open-ended approach, closed-ended approach or a payment scale method. An expert CV panel criticised the open-ended approach as producing "biased and erratic" results (Arrow et al., 1993). To overcome
some of these problems they suggested the use of a referendum type format with respondents raising or lowering bids with yes/no responses. However, they did not consider the use of payment scales. One of the other sources of bias is starting-point bias and evidence regarding this is mixed. O’Brien and Viramontes (1994) found no association between starting and final bids. However, Stålhammar (1996) found a statistically significant higher WTP when a higher starting bid was used.

A study by Donaldson et al. (1995b) used both open-ended and payment scale methods to investigate how much respondents would be willing to pay for bone scans to detect osteoporosis. They found that a payment scale method increased the response rate overall, gave a higher rate of completion of the WTP question, and gave higher mean and median WTP values. In addition, there were no zero values with the use of the payment scale. Obviously there is still range bias (bias generated by the selection of the values) for the closed-ended and payment scale methods and this is an area which requires further investigation (Donaldson et al., 1995b).

1.6.4 Studies involving the WTP approach

WTP in health care has been used in three main contexts (Morrison and Gyldmark, 1992):

- Valuing prevention. Persson and Cedervall (1991, cited in Morrison and Gyldmark, 1992) asked subjects’ maximum WTP values for safety devices in cars which would reduce risk of fatal injury. In addition, they also determined WTP to reduce the risk of contracting influenza, a bleeding ulcer and a fatal heart attack.

- Valuing treatments and services. Berwick and Weinstein (1985) used a direct open-ended valuation technique to determine how much women would be WTP for information in connection with ultrasound scans in normal pregnancy. Donaldson (1990b) asked relatives and friends of elderly residents of NHS nursing homes and continuing care hospital wards how much more they would be WTP to keep a relative or friend in the preferred choice of accommodation. They found that NHS
nursing homes were a more efficient option for the continuing care of the elderly if, for purposes of social planning, only one option was chosen.

- Valuing outcomes or health states. Relatively few studies have focused on health states *per se* rather than treatments. Thompson (1986) determined WTP for a cure for rheumatoid arthritis, finding that respondents were willing to pay 22% of their income for a cure. Although the question of a complete cure may seem unrealistic, it did allow the author to estimate the burden of arthritis in monetary terms.

O'Brien and Viramontes (1993, 1994) studied WTP in patients with chronic respiratory disease and found that the mean WTP for treatment was $113 per month but with a median of $65 and a standard deviation of $205. They concluded that WTP did not offer good discrimination as there was no association with severity of disease. However, it did offer useful indicators.

An investigation of the maximum WTP for a change in angina symptoms was undertaken by Chestnut et al. (1996). Regression analysis found that the strongest relationships existed between WTP and household income, previous bypass surgery, WTP starting point and current frequency of angina episodes. Estimates of WTP to avoid angina episodes were of the same general magnitude as estimates calculated from the patient records of actual expenditure and perceived episodes avoided.

Donaldson *et al.* (1996) investigated valuation of preventive techniques in health care, specifically whether WTP was a feasible method of estimating the benefits of avoiding food-borne risk. An estimate of consumers' valuations for the benefits of poultry irradiation were assessed by asking people how much extra they would be prepared to pay for poultry which had been treated by this preventive measure. Half of the subjects were willing to pay extra to reduce the risk of food-borne illness.
Olsen and Donaldson (1998) asked 150 interviewees to estimate their WTP in increased taxation for three services: an air ambulance; more heart operations and more hip replacements. They then compared QALYs gained from each programme with WTP values to determine whether the alternative methods resulted in different rankings. The most striking aspect of the comparison was the much lower valuation in terms of WTP for QALYs gained from hip replacements. The reason given most frequently was that "hips do not save lives". This led to the conclusion that QALYs gained from life improvements are valued lower than those which are life extending or life saving. The highest valued QALYs were for the helicopter ambulance where life saving was at a critical risk level.

Donaldson and Shackley (1997) investigated WTP for laparoscopic cholecystectomy to assess whether respondents would be willing to pay more for improvements in the process of care in addition to improvements in outcome. Their results were actually contrary to the hypothesis and showed that WTP was greater in the "outcome only" group than in the "process and outcome" group. This study was, therefore, not able to establish the existence of process utility and the authors reinforced the need for further investigations in this area.

1.7 ECONOMIC EVALUATION IN DENTISTRY

"When alternative therapies are available, patients want the choice of treatment to be based on processes that are cost-effective and have proven outcomes."

(Feine et al., 1998)

It is likely there will be an increased demand for economic analyses of dental interventions by the public and by those funding health care. Both the NHS and private companies are likely to demand evidence of value for money in the future. This is particularly important in areas which may be perceived as "cosmetic".
Economic evaluation is still used less frequently in dentistry than in medicine. However, this is beginning to change. A computerised Medline search in April 2000 showed the following:

- "cost-effectiveness" and "dentistry" produced 408 papers published between 1971 and 1999 with 78 in 1998/2000; "cost-effectiveness" and "orthodontics" produced 23 papers;
- "cost-benefit" and "dentistry" produced a total of 388 papers published between 1971 and 1999 with 42 of those in 1998/2000; "cost-benefit" and "orthodontics" produced 20 papers;
- "cost-utility" and "dentistry" produced only 19 papers, all of which were published between 1980 and 2000;
- "cost-utility" and "orthodontics", "cost-minimisation" and "dentistry" and "cost-minimisation" and "orthodontics" did not identify any papers.

It is, however, worth noting that a number of papers were listed under both cost-benefit and cost-effectiveness. This stresses the importance that papers must be read carefully to determine which method of analysis was actually used. It is also worth noting that a large number of the papers listed had not undertaken any form of economic analysis and merely mentioned that economic evaluation would be a useful next step in research, reflecting poor use of key words. A relatively small number of the papers had undertaken carefully controlled economic evaluation.

### 1.7.1 Cost-effectiveness and cost-benefit studies in dentistry

Cost-effectiveness and cost-benefit studies are therefore carried out much more frequently than cost-utility studies. This probably reflects the increased difficulty and time consuming nature of utility studies. However, the utility method is particularly useful in the field of dentistry because treatments frequently produce improvements in quality of life.
Cost-effectiveness and cost-benefit studies have focused largely on comparison of restorative materials (Mjör, 1992; Smales and Hawthorne, 1996; Mjör et al., 1997). Mjör (1992) studied the cost-effectiveness of restorative materials for two and three surface restorations undertaken in Norway and found amalgam to be the most cost-effective, followed by composite and then gold. A similar study in the UK confirmed these findings (Mjör et al., 1997). It was proposed that the cost-effectiveness of composites in particular was lower due to the shorter longevity and the higher cost of these restorations. A recent NHS review looked at cost-effectiveness of intra coronal dental restorations (NHS Centre for Review and Dissemination, 1999). This was a systematic review of 652 studies looking at relative longevity and cost-effectiveness, 253 of these showed the minimum core of data to be included. The bulletin strongly recommended further research into the cost-effectiveness of dental restorations, particularly working from the premise that 60% of restorations are replacement restorations. It was also noted that a large number of economic studies were of poor quality and did not provide sufficient information.

Another major focus is preventive techniques. Klock (1980) looked at CBA and CEA of a preventive programme (including oral hygiene, fluoride application and fissure sealants) and found that in spite of a reduction in caries activity, the programme was uneconomic compared with traditional dental care. In contrast, Morgan et al. (1998) assessed the cost-effectiveness of a preventive programme in two non-fluoridated regions of Australia and concluded that the introduction of a preventive programme was an efficient use of resources. They also stressed the need for systematic evaluation of a full range of dental prevention and treatment programmes.

Edwards et al. (1999) studied the cost, the effectiveness and the cost-effectiveness of removal and retention of asymptomatic third molars. Using a decision tree model of outcomes and data from 100 patients, the authors concluded that mandibular third molar retention was less costly to the NHS, more effective for the patient and more cost-effective to both parties than removal. However, if the likelihood of developing pericoronitis or caries increased substantially, removal was more cost-effective.
1.7.2 Utility studies in dentistry

There are relatively few utility or cost-utility studies in the field of dentistry. A paper in 1997 noted the importance of training more personnel in the techniques necessary to undertake utility analyses (Sendi et al., 1997). It is surprising that the QALY has been used rarely in dentistry. However, other measures of utility which have been used are:

- Quality adjusted tooth year (QATY).
  This measure was introduced as an improvement on DMF. It is, however, a measure of quality of tooth and not quality of life (Fyffe and Nuttall, 1995).

- Quality adjusted prosthesis year (QAPY).
  This was introduced by Jacobson et al. (1992) and is an outcome measure for reconstruction of totally edentulous patients. The comments related to the QATY also apply to the QAPY.

Krischer (1976) investigated the utility structure of decision making in the treatment of cleft lip and palate (CLP). Utilities were assessed from a questionnaire and both clinicians and families of CLP children were included. The author found significant differences between clinicians and families when assessing cosmetic outcome and speech. However, because the questionnaire was devised by the author and did not use one of the standard methods of utility assessment, it is difficult to draw conclusions.

Fyffe and Kay, in 1992, assessed utility values for four different “tooth states” which, it was hypothesised, would have different values. They found that the highest mean utility values were for the restored tooth and lowest values for the decayed and painful posterior tooth. Values were obtained from both dentists and the general public. Perhaps not surprisingly, dentists gave higher utility values when compared with members of the general public.
Armstrong et al. (1995) and Brickley et al. (1995) studied the relative utility values for possible outcomes of surgery and non-intervention in the management of third molars. They found that post-operative problems were considered to reduce health more than non-intervention associated with recurrent mild pericoronitis. A similar study by Bellamy et al. (1996) looked at multi-attribute utility theory for periodontal patients and developed a multi-attribute scale based on interviews with patients undergoing periodontal treatment.

Downer et al. (1997) used a convenience sample to elicit the public’s perceptions of different oral cancer states (precancer, small cancer and large cancer). They used a standard gamble questionnaire and found utility values of 0.92 for precancer, 0.88 for stage 1 cancer and 0.68 for stage 2 cancer. There were statistically significant differences between all three values with the order of magnitude in the expected direction.

Downer and Moles (1998) used a computer simulation to study the influence of relevant factors on health gain from restorative treatment under varying assumptions and compared this with a “do nothing” approach. A total of 92 adults were given a standard gamble questionnaire and asked to choose between having a posterior tooth in a certain state of health or disease versus a perfectly healthy tooth/extraction gamble. The results suggested that caution is advisable when making positive treatment decisions to restore teeth on the basis of bitewing radiographs. The authors also highlighted the urgent need for cost-utility analyses within the dental field.

O’Brien et al. (1998) undertook the only example which was found of utility analysis in orthodontics. They developed a TTO questionnaire using the aesthetic component of the Index of Treatment Need (Evans and Shaw, 1987) and found that patients seeking treatment gave lower utility values for the aesthetic components 5 and 8 than those not seeking treatment. In contrast, with the visual analogue scale there was no significant difference for 3, 5 or 8 on the aesthetic component. It, therefore, appeared
that seekers and non-seekers may have similar perceptions of malocclusion but place different utilities on those dental states.

Another current area of interest is that of implant retained prostheses. MacEntee and Walton (1998) looked at the costs associated with implant retained prostheses and conventional dentures. Jacobson et al. (1992) also compared implant retained prostheses and conventional dentures using the RS method. They concluded that this was a reliable measure of patients' preferences and the implant group rated a successful implant-supported prosthesis as higher than a functional, fitting, aesthetic conventional denture, in spite of higher costs and longer periods of non-function.

A recent example of clinical trials and economic evaluation being undertaken concurrently is that by Severens et al. (1998) who assessed the short-term cost effectiveness of pre-surgical orthopaedics in babies with a complete unilateral cleft of the lip and palate. There was a significant difference in both medical and indirect costs for the two groups with those for the pre-surgical orthopaedic group being higher. However, the outcome, which was assessed in terms of operating time, was found to be non-significant. Thus concluding that pre-surgical orthopaedics was not cost-effective in terms of reduced operating time. Other important outcome measures such as appearance and function are to be reported at a later stage.
1.8 SUMMARY

Economic evaluation is now an accepted method for appraisal of health care interventions. This may take one of several forms, with cost-utility and cost-benefit being particularly useful techniques. There are still only a small number of papers describing economic evaluation in dentistry but this is likely to change in future years as the NHS and private health companies ask for evidence that interventions they are funding are worthwhile and justified.
2.0 AIMS

2.1 UTILITY ANALYSIS

The aims of the utility analysis study were to:

• obtain utility values for the pre-treatment health state for patients requesting combined orthodontic/maxillofacial surgery to correct dentofacial abnormalities and to compare these with values obtained from a control group of non-patients.

• study a number of patients longitudinally, comparing utility values at various stages of treatment and to calculate the cost per QALY for orthognathic treatment.

• determine whether the process of orthognathic treatment is associated with loss of QALYs.

2.2 WILLINGNESS-TO-PAY

The aims of the willingness-to-pay study were to:

• elicit willingness-to-pay values from orthognathic patients and from a control group.

• determine whether willingness-to-pay truly measures strength of preference.
3.0 MATERIALS AND METHODS

Ethical approval was obtained from the Joint Research and Ethics Committee of the Eastman Dental Institute and Hospital. All subjects were provided with an information sheet and completed a consent form. All patients were assured of confidentiality and informed that failure to participate in the study or withdrawal from the study at any time would not affect their present or future treatment.

3.1 INSTRUMENTS

The method used in this study to obtain utility values was based on that recommended in the CHEPA Working Series #90-9 developed by McMaster University, Canada (Furlong et al., 1990). Three methods were used: the rating scale (RS), standard gamble (SG) and time trade-off (TTO).

One of the most important factors when asking members of the general public to provide utility values is that they must understand fully the health state being described. The study achieved this using both verbal and visual explanations. The problems associated with dentofacial deformity and the process of orthognathic treatment were described to each respondent (Appendix 1) and black and white photographic records of four patients were also shown (male and female; class II and class III malocclusions). These individuals had previously given permission for their photographic records to be used in publications. Respondents were instructed to consider that the health state applied to themselves (Sackett and Torrance, 1978; Torrance, 1982). The experimental group (patients) were asked to answer the questions with reference to their own health state.
3.1.1 Rating scale¹ (see Appendix 1)

The rating scale was a vertical and calibrated visual analogue scale (0-100), as described by Feeny and Torrance (1989). The scale was laser printed and laminated. Initially the respondents were asked to “anchor” the two ends of the scale. This was achieved using two health state descriptions:

- Perfect health state: respondents were “forced” to assign perfect health to a scale value of 100.
- Death: respondents were told that if death was the worst health state they could imagine, it should be placed at 0. However, if worse health states could be imagined (for example, being in constant pain), death was placed at an intermediate point (d) on the scale.

The scale was then used to record a preference value (x) for the health state in question.

3.1.2 Standard gamble (see Appendix 1)

Subjects were asked to choose between two sets of health-related personal circumstances. In the first situation (Choice A) they were told to envisage that they would live with the dentofacial deformity for the rest of their life. In the second situation (Choice B) they were asked to envisage a gamble situation where there were two possible outcomes: perfect health with no dentofacial deformity or death. The probability of perfect health was decreased incrementally and the probability of death increased incrementally. The value at which the subject was not willing to take any further risk of death was regarded as the indifference point and was recorded as the utility value. The more risk a respondent was prepared to accept, the lower the utility value.

¹ The RS generally uses a 0-100 rather than a 0-1 scale. The values are then converted to the standard 0-1 scale (by dividing by 100) for calculation of QALYs etc.
Figure 4 The standard gamble technique.

The principles involved in the standard gamble may be difficult for some respondents to understand and visual aids were used to make the concepts clearer. A card deck was constructed using laminated diagrams in which colour coded pie-chart segments represented the probabilities. In order to avoid an anchoring bias, where values were constantly increased or decreased, a “ping-pong” technique was used. This involved alternating values between high and low percentages (10/90; 90/10; 20/80; 80/20 etc.) and reduced the risk of respondents over or underestimating the indifference point.

The decision to use the perfect health/death approach is considered more fully in the discussion.

3.1.3 Time trade-off (see Appendix 1)

The study assumed the remaining life expectancy to be 50 years, as the mean age of the control group respondents was 25.1 years and that of the experimental group was 23.5 years. Subjects were offered Choice A (State i) which was 50 years with dentofacial deformity or Choice B which was a shorter life span but in perfect health. The time in the healthy state (time x) was varied until an indifference point was
reached, at which the utility value was recorded. As with the SG, bias was reduced by alternating between long and short periods (for example, 45 years, followed by 5, 40, 10, 35, 15, 30, 20, 25). The more years a respondent was prepared to forgo, the lower the utility value.

Figure 5 The time trade-off technique.

3.1.4 Willingness-to-pay

Willingness-to-pay values were ascertained using the payment card method. This method was selected as it has been suggested that the payment card method is more valid than an open-ended approach (Donaldson et al., 1995b). A series of cards were presented with values ranging from £0 to £15,000. These values were chosen as the range was acceptable and incorporated within it the actual cost of the procedure. At the time that this study was initiated, the extra contractual referral (ECR) cost for orthognathic surgery was £7,000. Therefore this was deemed an appropriate range as the actual cost was approximately in the middle. Respondents were asked how much they would be prepared to pay to undergo treatment for the correction of their dentofacial deformity. They were advised to consider the amount as being a once only payment but which could be paid in installments over several years if required (this was likened to a bank loan).
A number of the respondents were students and not financially self-sufficient. Greater than 50 percent of these individuals did not know their household income and it was therefore decided that it was not feasible to relate willingness-to-pay values to household income. However, in an attempt to investigate the influence of ability to pay, postcodes were used to derive a Townsend deprivation score for each respondent in the experimental group (Townsend et al., 1986; Gilthorpe et al., 1997). The Townsend index consists of four population variables: percentage measures of local unemployment; home ownership; overcrowding and car ownership. The index may have a negative or positive value as the mean for the population from which it is derived is zero. Positive values indicate greater deprivation and negative values show greater affluence. The relationship between this index and WTP was investigated.

3.2 SUBJECTS

3.2.1 Sample size

Comparison of experimental and control groups

Sample size calculation was undertaken using non-parametric options in nQuery Advisor® Release 3.0. The Mann-Whitney programme is based on an estimation of the differences between the two groups. However, as there was little previous data from which to draw useful information, assumptions regarding what magnitude of differences would be important were based on advice from colleagues and also on clinical judgement. It must be acknowledged that smaller differences were not discounted as unimportant.
The programme uses a common standard deviation (SD) from the two groups and SDs were obtained from the first 20 subjects (10 experimental and 10 control) who completed interviews. Although this data could also have been used to estimate the clinically important differences, it was felt that with only 20 subjects the differences in scores may be non-representative and misleading.

Table 7 Sample size calculations.

<table>
<thead>
<tr>
<th></th>
<th>Assumed clinically important difference (SD for first 20 cases)</th>
<th>Number of subjects required in each group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>80% Power</td>
</tr>
<tr>
<td>Rating scale</td>
<td>10 (16)</td>
<td>45</td>
</tr>
<tr>
<td>Standard gamble</td>
<td>0.1 (0.19)</td>
<td>63</td>
</tr>
<tr>
<td>Time trade-off</td>
<td>0.1 (0.21)</td>
<td>76</td>
</tr>
<tr>
<td>Willingness-to-pay</td>
<td>£2,500 (£5,218)</td>
<td>74</td>
</tr>
</tbody>
</table>

Longitudinal data (experimental group only)

nQuery Advisor® Release 3.0 was also used to establish the number of subjects required in the longitudinal study. As it was the change in utility as a result of treatment which was of primary concern, the utility values at T1 and T5 were the important values. Assumptions regarding what magnitude of differences would be important were judged as in section 3.3.1, with standard deviations of the differences between the T5/T1 values obtained from the first 8 subjects who completed all interviews.
Table 8  Sample size calculations (longitudinal study).

<table>
<thead>
<tr>
<th></th>
<th>Assumed clinically important difference (SD of T5/T1 difference for first 8 completed cases)</th>
<th>Number of subjects required</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>80% Power</td>
</tr>
<tr>
<td>Rating scale</td>
<td>15 (25)</td>
<td>24</td>
</tr>
<tr>
<td>Standard gamble</td>
<td>0.1 (0.11)</td>
<td>12</td>
</tr>
<tr>
<td>Time trade-off</td>
<td>0.1 (0.17)</td>
<td>25</td>
</tr>
<tr>
<td>Willingness-to-pay</td>
<td>£2,500 (£2,808)</td>
<td>12</td>
</tr>
</tbody>
</table>

3.2.2  Recruitment of subjects

Control Group
A convenience sample of 80 adults were recruited to take part in the study as the control group. Subjects included students from local colleges, staff at local businesses and non-clinical ancillary staff at the Eastman Dental Institute and Hospital. The control group was a separate group from those individuals who completed questionnaires for Chapter 4. All subjects were fit and healthy and none had any personal involvement in orthognathic procedures. All subjects were over 16 years of age and spoke fluent English.

Respondents in the control group were interviewed once. All interviews were by the same person (Miss S. Cunningham) and lasted approximately 30 minutes. Utility values for the pre-treatment health state for an individual requesting orthognathic treatment were obtained from all 80 respondents. In addition, the respondent’s WTP for treatment was also established.
Experimental group
Sixty patients from the Eastman Dental Hospital were recruited to form the experimental group. The following criteria had to be fulfilled for inclusion in the study:
- over 16 years of age;
- spoke fluent English;
- medically fit and well;
- the dentofacial deformity was not due to a cleft of the lip/palate or other congenital deformity.

Two subjects declined to take part in the study; one started the interview but was found to have an unrelated medical condition which could influence utility values and two were subsequently excluded as it was felt that their spoken English was not adequate. The experimental group therefore comprised 55 patients. The majority of experimental group respondents (37 individuals) were involved only in this section of the study, although 18 were also involved in the questionnaire section (Chapter 4). It was aimed to have minimal cross-over between the two groups in order to avoid overburdening certain patients. All interviews took place prior to any active treatment and away from the clinical area. Interviews were undertaken by Miss S. Cunningham and took approximately 30 minutes to complete.

Experimental group: longitudinal analysis
The first 25 patients recruited for the experimental group were selected to form a group for longitudinal analysis. Interviews were undertaken five times during treatment with a total interview time of approximately 2.5 to 3 hours for each patient:
- T1 Prior to starting treatment;
- T2 Pre-op: 6 to 8 months into pre-surgical orthodontics;
- T3 Pre-op: 12 to 14 months into pre-surgical orthodontics;
- T4 Post-op: approximately 2 weeks following wafer removal/ 4 weeks post-op;
- T5 2 to 4 weeks following debond.
Unfortunately, due to cancellation of operations in late 1999 and early 2000, three individuals had not completed treatment at the time of data analysis. In addition, one respondent was unable to complete the T5 interview at the appropriate time due to illness. Therefore, complete data was available for 21 of these 25 patients.

3.3 REPEATABILITY OF THE METHOD

In order to establish the repeatability of each of the three utility methods and the WTP method, twenty five control group respondents underwent a repeat interview four to six weeks after the first.

3.4 CALCULATION OF QALYs GAINED (Longitudinal group)

Total QALYs gained were calculated individually for each respondent and each of the three methods, using the following equation (in all cases, the life expectancy was assumed to be 75 years):

\[ \text{QALYs gained} = \frac{\text{change in utility value due to treatment}}{\text{number of years this benefit exists}} \times \text{number of years this benefit exists} \]

Three plots were drawn for each respondent and examples of these are shown in Appendix 2. Certain criteria had to be established for calculating the QALYs lost or gained during treatment. One of these was that there was a linear relationship between consecutive utility measurements. This created some difficulty for the T3/T4 time span as it was felt likely that any increases (or decreases) in value would be associated with the surgical intervention rather than any influences at T3 itself. Therefore, an assumption was made that the T3 value would continue until surgery and then any increases (or decreases) would occur between this time point and T4. This is
illustrated in the example shown below. The area of each section of the graph (labelled 1 to 5) was calculated using the formula shown below.

\[ QALYs \text{ gained or lost} = \]
\[ \text{change in utility value due to treatment} \times \text{number of years (or part of) that this benefit/loss exists} \]

**Figure 6** QALYs lost/gained during treatment.

Utility value

\[ T1 \leftarrow 6-8 \text{ months} \rightarrow T2 \leftarrow 6-8 \text{ months} \rightarrow T3 \leftarrow \text{Variable} \rightarrow \leftarrow 2-4 \text{ wks} \rightarrow T4 \leftarrow \text{Variable} \rightarrow T5 \]

Surgery

NB: Area under the initial T1 value represents QALYs lost during treatment.
Area above initial T1 value represents QALYs gained during treatment.

\[ \text{Total QALYs gained} = \]
\[ [\text{area of (1+2+3+4+5)}] + [(T5 \text{ utility}-T1 \text{ utility}) \times \text{Future life expectancy}] \]

For areas 1 to 5, QALY calculations were based on years (or part of a year).
All changes in utility value were taken with respect to the initial pre-treatment value:

- **Area 1**
  This represents the change over the T1/T2 period. The utility value is assumed to reduce as soon as the fixed appliances are placed.

- **Area 2**
  This represents the change over the T2/T3 period. The utility value is still below the T1 value but has increased a little, presumably as the patient becomes used to treatment.

- **Area 3**
  This represents the change between T3 and surgery, assuming the situation remains relatively stable until the surgical intervention.

- **Area 4**
  This represents the change between surgery and T4. There is a gradual increase in utility as the patient recovers from surgery.

- **Area 5**
  This represents the change in utility over the time period T4/T5. The utility value increases beyond T1 levels, assuming a gradual change in the final stages of treatment.
Figure 7  Calculation of total QALYs gained.

Utility value

QALYs = Area of (1+2+3+4+5) + Area of 6
3.5 ESTIMATION OF COSTS

The resources used were calculated for each of the individuals involved in the longitudinal study. These are shown in Appendix 3 and are divided into:

- Orthodontics;
- Miscellaneous (i.e. hygienist visits, extractions by GDPs, restorative care);
- Out-patient;
- In-patient;
- Operating theatre.

Resource use was from the perspective of the NHS and, as such, did not include costs accrued by the patient or their family. Nor was there any consideration of the capital costs of building space. Costs were adjusted to the final year of the study i.e. April 1999 to April 2000.

3.6 SENSITIVITY ANALYSES

A number of sensitivity analyses were conducted:

3.6.1 Orthodontic costs

The base cost for each patient for orthodontic treatment was calculated using the classic “bottom up” approach based on the number and duration of visits; the personnel involved at each visit; the exact sequence of archwires used and other consumables. The relevant information was derived from the hospital patient administration system (PAS), clinicians’ diaries and patients’ notes. This value formed the base estimate and was also the lowest of the three estimates used in the sensitivity analysis. The alternative to this method of costing involved using existing costs per visit. A cost per visit (£28-00) was provided by University College London Hospitals (UCLH) Trust based on that quoted to purchasers. However, this is subsided due to the hospital being a teaching institute so an alternative cost per visit was also used.
(£42-00) based on the 1998 cost of an out-patient visit to dental specialties (Chartered Institute of Public Finance and Accountancy, 1998). These represented a mid and high estimate respectively.

3.6.2 Cost of each night stay in hospital

The cost per night for each patient is virtually impossible to calculate, therefore *per diem* costs were used. A range of values were introduced into the sensitivity analysis. The quoted value for UCLH Trust was £150/night (introduced as the base cost), which was comparable with other studies (Sculpher, Personal Communication). A low and high estimate of £100/night and £200/night respectively were also used.

3.6.3 Theatre cost

Theatre costs were calculated using a “bottom up” approach based on the cost per minute to run the operating theatre (which includes only overheads and non-clinical staff) plus the cost of all staff and consumables. The cost per minute was calculated by UCLH Trust accountants as £0.96 and this was comparable with that quoted in other publications when inflation was taken into account (Bevan, 1989; Sculpher *et al.*, 1993; NHS Executive, 1999). Staff costs were judged from the personnel present in theatre for individual patients, as obtained from the operation notes. Costs for consumables were estimated based on known use of certain consumables for all patients and titanium plates and screws were calculated for each individual respondent. Use of drugs and anaesthetic agents was judged from individual drug/anaesthetic charts. This calculation (Method 1) was accepted as the base estimate as it was felt to be the most accurate method. It was also included in the low cost estimate.

UCLH Trust also provided a cost per theatre session of £1947-00 (excluding surgeons and special consumables i.e. titanium plates). However, this was considered to be higher than average so this cost was included only in the high cost estimate (Method 2).
CHAPTER 2 MATERIALS AND METHODS

Three patients underwent additional related surgical procedures in the day stay unit (including removal of plates and closure of a palatal fistula) therefore costs for these procedures were calculated separately (Appendix 3).

3.6.4 Miscellaneous

Two patients involved in this study also required complex restorative treatment including dental implants. This treatment may have been undertaken, however, whether the individuals had orthognathic treatment or not. Therefore, the costs incurred were included in the sensitivity analysis but only in the high cost estimate.

3.6.5 Summary of resource use

Base cost estimate comprised:
• cost of orthodontic treatment as calculated (Appendix 3);
• extractions;
• out-patient costs as calculated;
• in-patient costs (£150 per night);
• theatre costs as calculated in method 1 (Appendix 3).

Low cost estimate comprised:
• cost of orthodontic treatment as calculated (Appendix 3);
• extractions;
• out-patient costs as calculated;
• in-patient costs (£100 per night);
• theatre costs as calculated in method 1 (Appendix 3).

High cost estimate comprised:
• cost of orthodontic treatment based on £42.00 per visit;
• extractions;
• restorative treatment required as part of the overall treatment;
• out-patient costs as calculated;
• in-patient costs (£200 per night);
• theatre costs as calculated in method 2 (Appendix 3).

NB: Outpatient costs were not subjected to sensitivity analysis as the unique nature of the Joint Clinics (with several clinicians present) did not lend itself to cost per visit analysis as in Section 3.6.1.

3.7 DISCOUNTING COSTS AND BENEFITS

In this study, benefits (QALYs) were discounted at 2% and costs at 6%. QALYs in treatment between T1 and T2 were discounted to T1, those gained between T2 and T3 were discounted to T2 etc. using the equation shown below. This is a standard formula applied to discounting rates per annum but in this instance was also used for periods shorter than a year. The mid-time point was used for discounting in each case with the time period in the example below being 6 months (midway=3 months or 0.25 year).

\[
QALYs \text{ (T1-T2)} \times \frac{1}{(1 + d)^{0.25}}
\]

where \( d = \text{discount rate (0.02 or 2\% for QALYs and 6\% or 0.06 for costs)} \)

Total QALYs gained as a result of treatment were discounted assuming that the benefits occurred at the beginning of each of the future years. Therefore, the benefits for year 1 did not need to be discounted but those for each of the following years were discounted at the appropriate rate.

It was not possible to know at exactly which time point each of the costs were incurred, therefore, costs were discounted midway between the start of treatment and surgery, based on the fact that the majority of the costs occurred during this time period. The same equation as above was used.
3.8 BOOTSTRAPPING OF RESULTS

The non-parametric bootstrap technique was applied to the total QALY and resource use data (Efron and Tibshirani, 1993). Only the data for the bimaxillary and single jaw groups combined were analysed in this way as it was deemed inappropriate to bootstrap data for groups any smaller than this. The number of replications in each case was 2000.
3.9 STATISTICAL SOFTWARE AND PROCEDURES

Data were analysed using the SPSS for Windows\(^1\) package. Comparison of utility values between experimental and control groups used the Mann-Whitney U test and comparison of values within groups used the Wilcoxon signed-rank test. Spearman correlation coefficients were calculated for the relationships between age and utility values; WTP and utility values; and WTP and Townsend deprivation scores. Test-retest reliability used the methodology described in Chapter 1. Method agreement (reproducibility) used the Bland and Altman method (1986). 95% confidence intervals were checked using the Simcalc\(^2\) programme. Bootstrap calculations were undertaken using the SAS\(^3\) programme.

---

\(^1\) SPSS for Windows. SPSS UK Ltd. 1st Floor St. Andrew's House, West Street, Woking, Surrey, GU21 1EP.

\(^2\) SIMATH-Gruppe, Lehrstuhl Prof. Dr. H.G. Zimmer, FB9 Mathematik, Universitaet des Saarlands, Postfach 151150, D-66041, Saarbruecken, Germany.

4.0 UTILITY ANALYSIS RESULTS

4.1 COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS

4.1.1 Demographic details

As anticipated for the experimental group, the ratio of females to males was approximately 2:1. This reflected the ratio of females to males seeking orthognathic treatment. In view of this, a similar ratio was recruited for the control group.

It was also aimed to have a similar age range in both groups (16 to 42 and 16 to 43 years) with a resultant mean age of 23.5 years for the experimental group and 25.1 years for the control group.

Table 9 Details of respondents.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (years)</th>
<th>Age range (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female</td>
<td>Mean (95% CI)</td>
</tr>
<tr>
<td>Experimental</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>Control</td>
<td>24</td>
<td>56</td>
</tr>
</tbody>
</table>
4.1.2 Reliability

Reliability of all three utility methods are shown in Table 10.

Table 10 Test-retest data.

<table>
<thead>
<tr>
<th>Method</th>
<th>Pearson correlation</th>
<th>ICC</th>
<th>SD of differences</th>
<th>Repeatability coefficient (1.96 x SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating scale (Range 0-100)</td>
<td>0.64**</td>
<td>0.64</td>
<td>14.40</td>
<td>0.27</td>
</tr>
<tr>
<td>Standard gamble (Range 0-1)</td>
<td>0.76**</td>
<td>0.74</td>
<td>0.10</td>
<td>0.20</td>
</tr>
<tr>
<td>Time trade-off (Range 0-1)</td>
<td>0.66**</td>
<td>0.66</td>
<td>0.13</td>
<td>0.25</td>
</tr>
</tbody>
</table>

**p<0.01

4.1.3 Utility values and method agreement

The utility values obtained using all three methods for both experimental and control groups are shown in Table 11. Table 12 presents the results of the Mann-Whitney test comparing the three methods between the two groups. This confirmed significant findings between the experimental and control groups for the RS (p=0.02) but found no significant differences for the SG or TTO. Table 13 reports no significant correlations between the utility values and age for either group.

Method agreement comparing the RS, SG and TTO is shown in Table 14.
Table 11  
Comparison of utility values for both groups using all three methods.

<table>
<thead>
<tr>
<th>Group</th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>63.49 (19.75)</td>
<td>0.76 (0.21)</td>
<td>0.68 (0.25)</td>
</tr>
<tr>
<td>95% CI</td>
<td>58.15 to 68.83</td>
<td>0.70 to 0.82</td>
<td>0.61 to 0.75</td>
</tr>
<tr>
<td>Median</td>
<td>60.0</td>
<td>0.85</td>
<td>0.75</td>
</tr>
<tr>
<td>Range</td>
<td>10-98</td>
<td>0.05-0.95</td>
<td>0.05-0.95</td>
</tr>
<tr>
<td>Control</td>
<td>55.38 (17.91)</td>
<td>0.83 (0.12)</td>
<td>0.74 (0.15)</td>
</tr>
<tr>
<td>95% CI</td>
<td>51.39 to 59.37</td>
<td>0.80 to 0.86</td>
<td>0.71 to 0.77</td>
</tr>
<tr>
<td>Median</td>
<td>58.50</td>
<td>0.85</td>
<td>0.75</td>
</tr>
<tr>
<td>Range</td>
<td>11 - 94</td>
<td>0.45 - 0.95</td>
<td>0.25 - 0.95</td>
</tr>
</tbody>
</table>

NB: No significant difference between males and females.
Table 12  Comparison of mean utility values between the experimental and control groups.

<table>
<thead>
<tr>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>p-value</td>
<td>U</td>
</tr>
<tr>
<td>1679.5</td>
<td>0.02*</td>
<td>1919.5</td>
</tr>
</tbody>
</table>

U = Mann-Whitney U value  
* p<0.05

Table 13  Spearman correlations between age and utility values for the experimental and control groups.

<table>
<thead>
<tr>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>p-value</td>
</tr>
<tr>
<td>Experimental</td>
<td>0.25</td>
<td>0.06</td>
</tr>
<tr>
<td>Control</td>
<td>-0.01</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Table 14  Method agreement between the three utility measurement methods.

<table>
<thead>
<tr>
<th></th>
<th>Pearson correlation</th>
<th>Bland and Altman analysis of method agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SG/TTO</td>
<td>SG/RS</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=55)</td>
<td>0.75**</td>
<td>0.46**</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=80)</td>
<td>0.38**</td>
<td>0.35**</td>
</tr>
<tr>
<td>Groups combined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=135)</td>
<td>0.63**</td>
<td>0.35**</td>
</tr>
</tbody>
</table>

*p<0.05; **p<0.01
4.2 LONGITUDINAL STUDY OF UTILITY VALUES

T1 Prior to starting treatment
T2 Pre-op: 6-8 months into pre-surgical orthodontics
T3 Pre-op: 12-14 months into pre-surgical orthodontics
T4 Post-op: 2 weeks after wafer removal/ 4 weeks post-op
T5 2-4 weeks following debond

4.2.1 Demographic details

The longitudinal study group comprised 21 subjects, 15 of whom were female and 6 male. The mean age was 23.05 years (95% CI 19.96 to 26.14 years). Fifteen underwent bimaxillary surgery and 6 single jaw surgery (5 mandibular procedures and 1 maxillary procedure). Mean treatment duration was 23 months (95% CI 20.95 to 25.05 months).

Table 15 Details of respondents.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (years)</th>
<th>Surgical procedure</th>
<th>Duration of treatment (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female</td>
<td>Mean (95% CI)</td>
<td>Bimax.</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>23.05 (19.96 to 26.14)</td>
<td>15</td>
</tr>
</tbody>
</table>
4.2.2 Summary of utility values

Figures 8 to 10 represent summary plots for the mean utility values for all 21 subjects for the three methods. A similar pattern of changes is noted for the SG and TTO, with a markedly different pattern for the RS. Examples of individual utility plots are shown in Appendix 2.

Figure 8 Summary plot for all 21 cases (RS method).

Figure 9 Summary plot for all 21 cases (SG method).

NB: T3 utility value assumed to last until surgery and then increase to T4 levels
Table 16 presents the utility values for each of the three methods at each time point (T1 to T5). No significant differences were found between the bimaxillary or single jaw groups at T1 or T5 (Table 17). Table 18 shows significant differences between T1 and T5 utility values for the bimaxillary and single jaw group combined and for the bimaxillary group. However, no significant change in utility values was seen between T1 and T5 for the single jaw group.
Table 16  Utility values for all three methods at all five time periods.

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Median</td>
<td>Mean (SD)</td>
<td>Median</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Rating scale</td>
<td>58.24 (17.15)</td>
<td>50.00</td>
<td>70.00 (13.91)</td>
<td>74.00</td>
<td>75.14 (15.76)</td>
</tr>
<tr>
<td>Standard gamble</td>
<td>0.77 (0.22)</td>
<td>0.85</td>
<td>0.70 (0.23)</td>
<td>0.75</td>
<td>0.70 (0.18)</td>
</tr>
<tr>
<td>Time trade-off</td>
<td>0.72 (0.24)</td>
<td>0.75</td>
<td>0.67 (0.25)</td>
<td>0.75</td>
<td>0.64 (0.23)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Range</th>
<th>Range</th>
<th>Range</th>
<th>Range</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating scale</td>
<td>30-88</td>
<td>40-88</td>
<td>38-91</td>
<td>39-96</td>
<td>47-99</td>
</tr>
<tr>
<td>Standard gamble</td>
<td>0.05-0.95</td>
<td>0.05-0.95</td>
<td>0.25-0.95</td>
<td>0.55-0.95</td>
<td>0.75-1.00</td>
</tr>
<tr>
<td>Time trade-off</td>
<td>0.05-0.95</td>
<td>0.05-0.95</td>
<td>0.05-0.95</td>
<td>0.15-0.95</td>
<td>0.35-1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>95% Confidence Intervals</th>
<th>95% Confidence Intervals</th>
<th>95% Confidence Intervals</th>
<th>95% Confidence Intervals</th>
<th>95% Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating scale</td>
<td>50.43 to 66.05</td>
<td>63.67 to 76.33</td>
<td>67.97 to 82.31</td>
<td>66.38 to 83.62</td>
<td>78.58 to 90.52</td>
</tr>
<tr>
<td>Standard gamble</td>
<td>0.67 to 0.87</td>
<td>0.60 to 0.81</td>
<td>0.62 to 0.78</td>
<td>0.72 to 0.84</td>
<td>0.90 to 0.98</td>
</tr>
<tr>
<td>Time trade-off</td>
<td>0.61 to 0.83</td>
<td>0.56 to 0.78</td>
<td>0.54 to 0.75</td>
<td>0.61 to 0.82</td>
<td>0.83 to 0.97</td>
</tr>
</tbody>
</table>
### Table 17  Comparison of bimaxillary and single jaw group values at T1 and T5.

<table>
<thead>
<tr>
<th></th>
<th>RS T1</th>
<th>RS T5</th>
<th>SG T1</th>
<th>SG T5</th>
<th>TTO T1</th>
<th>TTO T5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U value</strong></td>
<td>41.50</td>
<td>33.50</td>
<td>35.00</td>
<td>32.50</td>
<td>42.00</td>
<td>31.00</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.79</td>
<td>0.38</td>
<td>0.47</td>
<td>0.34</td>
<td>0.85</td>
<td>0.30</td>
</tr>
</tbody>
</table>

$U =$ Mann-Whitney $U$ value

### Table 18  T1/T5 change: $Z$ values (and p-values).

<table>
<thead>
<tr>
<th></th>
<th>RS</th>
<th>SG</th>
<th>TTO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bimax and single jaw combined</strong></td>
<td>-3.66 (&lt;0.001)</td>
<td>-2.96 (0.003)</td>
<td>-3.23 (0.001)</td>
</tr>
<tr>
<td><strong>Bimaxillary surgery</strong></td>
<td>-3.23 (0.001)</td>
<td>-2.81 (0.005)</td>
<td>-3.07 (0.002)</td>
</tr>
<tr>
<td><strong>Single jaw surgery</strong></td>
<td>-1.57 (0.12)</td>
<td>-1.63 (0.10)</td>
<td>-1.24 (0.22)</td>
</tr>
</tbody>
</table>

$Z =$ Wilcoxon signed-rank test $Z$ value
4.2.3 QALYs gained

Tables 19 to 22 show mean QALYs, both non-discounted and discounted, for all three methods. Data are presented for bimaxillary and single jaw surgery groups combined and for both groups separately. The lower three rows of the tables present summary data for:

- QALYs gained or lost during treatment - variable (1)
- QALYs gained as a result of treatment - variable (2)
- Total QALYs gained - variable (1+2)

Figures 11 to 13 present the findings for the TTO with non-discounted and discounted mean values and 95% CI for variables 1, 2 and (1+2). The values shown are those derived from conventional analysis rather than the bootstrap method. Graphs for the RS and SG are very similar but are not included due to space constraints.

Table 23 compares the 95% CI for total QALYs gained using the conventional analysis and the bootstrap method. The values are similar in all cases with a slightly narrower range of 95% confidence intervals from the bootstrap method.
Table 19  Mean QALYs (95% CI) for all three methods for bimaxillary and single jaw surgery groups combined.

<table>
<thead>
<tr>
<th></th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean non-discounted value</td>
<td>Mean discounted value</td>
<td>Mean non-discounted value</td>
</tr>
<tr>
<td><strong>T1-T2</strong></td>
<td>0.03 (0.01 to 0.05 )</td>
<td>0.03 (0.01 to 0.05 )</td>
<td>-0.05 (-0.11 to 0.01 )</td>
</tr>
<tr>
<td><strong>T2-T3</strong></td>
<td>0.08 (0.04 to 0.12 )</td>
<td>0.08 (0.04 to 0.12 )</td>
<td>-0.04 (-0.10 to 0.02 )</td>
</tr>
<tr>
<td><strong>T3-Surgery</strong></td>
<td>0.07 (0.02 to 0.12 )</td>
<td>0.07 (0.02 to 0.12 )</td>
<td>-0.03 (-0.08 to 0.02 )</td>
</tr>
<tr>
<td><strong>Surgery-T4</strong></td>
<td>0.02 (0.01 to 0.03 )</td>
<td>0.02 (0.01 to 0.03 )</td>
<td>-0.01 (-0.02 to 0.00 )</td>
</tr>
<tr>
<td><strong>T4-T5</strong></td>
<td>0.08 (0.04 to 0.12 )</td>
<td>0.08 (0.04 to 0.12 )</td>
<td>0.03 (0.01 to 0.05 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>QALYs gained/lost during treatment (1)</th>
<th>QALYs gained as result of treatment (2)</th>
<th>Total QALYs (1+2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean discounted value</td>
<td>Mean discounted value</td>
<td>Mean discounted value</td>
</tr>
<tr>
<td><strong>T1-T2</strong></td>
<td>0.27 (0.15 to 0.39 )</td>
<td>0.27 (0.15 to 0.39 )</td>
<td>-0.10 (-0.27 to 0.07 )</td>
</tr>
<tr>
<td><strong>T2-T3</strong></td>
<td>13.75 (8.32 to 19.18 )</td>
<td>8.71 (5.30 to 12.12 )</td>
<td>8.23 (3.36 to 13.10 )</td>
</tr>
<tr>
<td><strong>T3-Surgery</strong></td>
<td>14.02 (8.50 to 19.54 )</td>
<td>8.98 (5.48 to 12.48 )</td>
<td>8.13 (3.14 to 13.12 )</td>
</tr>
</tbody>
</table>
Table 20  Mean QALYs (95% CI) for all three methods for the bimaxillary surgery group.

<table>
<thead>
<tr>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean non-</td>
<td>Mean discounted</td>
</tr>
<tr>
<td></td>
<td>discounted value</td>
<td>value</td>
</tr>
<tr>
<td>T1-T2</td>
<td>0.03 (0.01 to 0.05)</td>
<td>0.03 (0.01 to 0.05)</td>
</tr>
<tr>
<td>T2-T3</td>
<td>0.08 (0.03 to 0.13)</td>
<td>0.08 (0.03 to 0.13)</td>
</tr>
<tr>
<td>T3-Surgery</td>
<td>0.05 (0.01 to 0.09)</td>
<td>0.05 (0.01 to 0.09)</td>
</tr>
<tr>
<td>Surgery-T4</td>
<td>0.02 (0.00 to 0.04)</td>
<td>0.02 (0.00 to 0.04)</td>
</tr>
<tr>
<td>T4-T5</td>
<td>0.07 (0.02 to 0.12)</td>
<td>0.07 (0.02 to 0.12)</td>
</tr>
<tr>
<td>QALYs gained/lost</td>
<td>0.26 (0.13 to 0.40)</td>
<td>0.26 (0.13 to 0.40)</td>
</tr>
<tr>
<td>QALYs gained as</td>
<td>14.41 (7.91 to 20.91)</td>
<td>9.09 (5.09 to 13.09)</td>
</tr>
<tr>
<td>result of treatment</td>
<td>14.67 (8.08 to 21.27)</td>
<td>9.35 (5.26 to 13.44)</td>
</tr>
<tr>
<td>Total QALYs</td>
<td>14.67 (8.08 to 21.27)</td>
<td>9.35 (5.26 to 13.44)</td>
</tr>
</tbody>
</table>
Table 21  Mean QALYs (95% CI) for all three methods for the single jaw surgery group.

<table>
<thead>
<tr>
<th></th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean non-discounted value</td>
<td>Mean discounted value</td>
<td>Mean non-discounted value</td>
</tr>
<tr>
<td>T1-T2</td>
<td>0.01 (-0.03 to 0.05)</td>
<td>0.01 (-0.03 to 0.05)</td>
<td>-0.01 (-0.04 to 0.02)</td>
</tr>
<tr>
<td>T2-T3</td>
<td>0.06 (-0.02 to 0.14)</td>
<td>0.06 (-0.02 to 0.14)</td>
<td>-0.01 (-0.05 to 0.03)</td>
</tr>
<tr>
<td>T3-Surgery</td>
<td>0.13 (0.00 to 0.26)</td>
<td>0.13 (0.00 to 0.26)</td>
<td>-0.05 (-0.14 to 0.04)</td>
</tr>
<tr>
<td>Surgery-T4</td>
<td>0.01 (-0.04 to 0.06)</td>
<td>0.01 (-0.04 to 0.06)</td>
<td>-0.01 (-0.03 to 0.01)</td>
</tr>
<tr>
<td>T4-T5</td>
<td>0.08 (-0.01 to 0.17)</td>
<td>0.08 (-0.01 to 0.17)</td>
<td>0.04 (0.00 to 0.08)</td>
</tr>
<tr>
<td>QALYs gained/lost during treatment</td>
<td>0.30 (-0.03 to 0.63)</td>
<td>0.30 (-0.03 to 0.63)</td>
<td>-0.05 (-0.24 to 0.14)</td>
</tr>
<tr>
<td>QALYs gained as result of treatment</td>
<td>12.09 (-1.91 to 26.09)</td>
<td>7.77 (-1.42 to 16.96)</td>
<td>8.98 (-3.47 to 21.43)</td>
</tr>
<tr>
<td>Total QALYs</td>
<td>12.39 (-1.93 to 26.71)</td>
<td>8.07 (-1.43 to 17.57)</td>
<td>8.94 (-3.64 to 21.52)</td>
</tr>
</tbody>
</table>
Table 22  Mean total QALYs (95% CI).

<table>
<thead>
<tr>
<th></th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean non-discounted value</td>
<td>Mean discounted value</td>
<td>Mean non-discounted value</td>
</tr>
<tr>
<td>Groups combined</td>
<td>14.02 (8.50 to 19.54)</td>
<td>8.98 (5.48 to 12.48)</td>
<td>8.13 (3.14 to 13.12)</td>
</tr>
<tr>
<td>Bimaxillary</td>
<td>14.67 (8.08 to 21.27)</td>
<td>9.35 (5.26 to 13.44)</td>
<td>7.81 (1.75 to 13.87)</td>
</tr>
<tr>
<td>Single jaw surgery</td>
<td>12.39 (-1.93 to 26.71)</td>
<td>8.07 (-1.43 to 17.57)</td>
<td>8.94 (-3.64 to 21.52)</td>
</tr>
</tbody>
</table>
Figures 11-13  Examples of mean non-discounted and discounted QALYs using the TTO method (error bars represent 95% CI).

Bimaxillary and single jaw surgery groups combined

Bimaxillary surgery
Single jaw surgery

![Diagram showing QALYs (95% CI) for single jaw surgery. The x-axis represents different categories: QALYs in Rx, QALYs due to Rx, Total QALYs, QALYs in Rx (disc), QALYs due to Rx (dis), and Total QALYs (disc). The y-axis shows the QALYs values ranging from -10 to 30.]
Table 23  Comparison of 95% CI for total QALYs gained using conventional and bootstrap methods (bimaxillary and single jaw surgery groups combined).

<table>
<thead>
<tr>
<th></th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-discounted</td>
<td>Discounted</td>
<td>Non-discounted</td>
</tr>
<tr>
<td>Mean total QALYs</td>
<td>14.02</td>
<td>8.98</td>
<td>8.13</td>
</tr>
<tr>
<td>(conventional analysis)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(conventional analysis)</td>
<td>8.50 to 19.54</td>
<td>5.48 to 12.48</td>
<td>3.14 to 13.12</td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(bootstrap method)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.96 to 19.44</td>
<td>5.61 to 12.29</td>
<td>4.38 to 13.62</td>
</tr>
</tbody>
</table>
**Sensitivity analysis of QALYs**

Figure 14 and Table 24 represent a sensitivity analysis assuming that the utility value achieved at T5 is not retained throughout life, instead returning to T1 levels when the patient becomes older.

**Figure 14**  
Sensitivity analysis of QALYs gained: assuming T5 values are not retained for life (groups combined).
Table 24  Sensitivity analysis of total QALYs gained: assuming T5 values are not retained for life: mean values (95% CI).

<table>
<thead>
<tr>
<th>T5 utility value</th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retained until</td>
<td>Non-discounted</td>
<td>Discounted</td>
<td>Non-discounted</td>
</tr>
<tr>
<td>(years of age)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>14.02 (8.50 to 19.54)</td>
<td>8.98 (5.48 to 12.48)</td>
<td>8.13 (3.14 to 13.12)</td>
</tr>
<tr>
<td>65</td>
<td>11.32 (6.84 to 15.80)</td>
<td>7.88 (4.80 to 10.96)</td>
<td>6.51 (2.46 to 10.56)</td>
</tr>
<tr>
<td>55</td>
<td>8.62 (5.16 to 12.08)</td>
<td>6.53 (3.95 to 9.11)</td>
<td>4.89 (1.77 to 8.01)</td>
</tr>
<tr>
<td>45</td>
<td>5.92 (3.45 to 8.39)</td>
<td>4.88 (2.89 to 6.87)</td>
<td>3.27 (1.06 to 5.48)</td>
</tr>
</tbody>
</table>
4.2.4 Resource use

Appendix 3 presents the resource use for each individual stage of treatment. Sensitivity analyses of these data are presented in the following tables and figures. All figures show mean values, with 95% CI represented by error bars.

Table 25 and Figures 15 to 17 represent the sensitivity analysis of orthodontic treatment costs. Mean values ranged from a minimum value of £700.62 (for the bimaxillary group; base estimate) to a maximum value of £1,393.00 (for the single jaw group; high estimate).

Table 26 and Figures 18 to 20 represent the sensitivity analysis of in-patient hotel costs with mean values ranging from a minimum cost of £266.67 (for the single jaw group; low estimate) to a maximum cost of £760.00 (for the bimaxillary group; high estimate).

Data for the sensitivity analysis of operating theatre costs are shown in Table 27 and Figures 21 to 23. Mean values varied between a minimum of £997.83 (for the single jaw group; base estimate) to a high cost of £2,838.10 (for the bimaxillary group; high estimate). The sensitivity analysis of total resource use is presented in Table 28 and Figures 24 to 26. Resource use varied between a minimum of £2,549.79 (for the single jaw group; low estimate) to a maximum of £5,506.90 (for the bimaxillary surgery group; high estimate).

Table 29 presents data for total resource use in terms of non-discounted and discounted costs. Figures 27 to 29 show these findings graphically. Table 30 presents 95% confidence intervals for the total costs as calculated by conventional and bootstrap methods. The 95% CI for both methods were very similar.
Table 25  A sensitivity analysis of orthodontic treatment costs in response to alternative assumptions about costs incurred.

<table>
<thead>
<tr>
<th></th>
<th>Base estimate</th>
<th>@ £28.00 per visit</th>
<th>@ £42.00 per visit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
</tr>
<tr>
<td>Bimaxillary and single jaw combined</td>
<td>£707.99</td>
<td>£682.11 to £733.87</td>
<td>£841.33</td>
</tr>
<tr>
<td>Bimaxillary</td>
<td>£700.62</td>
<td>£671.44 to £729.80</td>
<td>£806.40</td>
</tr>
<tr>
<td>Single jaw</td>
<td>£726.41</td>
<td>£655.33 to £797.49</td>
<td>£928.67</td>
</tr>
</tbody>
</table>
Figures 15 to 17 Sensitivity analysis of orthodontic treatment costs: in £ Sterling (error bars represent 95% CI).

Bimaxillary and single jaw surgery groups combined

Bimaxillary surgery
Single jaw surgery
Table 26  A sensitivity analysis of in-patient hotel costs in response to alternative assumptions about cost per night.

<table>
<thead>
<tr>
<th></th>
<th>Base estimate (£150/night)</th>
<th>Low estimate (£100/night)</th>
<th>High estimate (£200/night)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
</tr>
<tr>
<td>Bimaxillary and single jaw combined</td>
<td>£521.43</td>
<td>£458.04 to £584.82</td>
<td>£347.62</td>
</tr>
<tr>
<td>Bimaxillary</td>
<td>£570.00</td>
<td>£498.41 to £641.59</td>
<td>£380.00</td>
</tr>
<tr>
<td>Single jaw</td>
<td>£400.00</td>
<td>£318.71 to £481.29</td>
<td>£266.67</td>
</tr>
</tbody>
</table>
**Figures 18 to 20** Sensitivity analysis of in-patient hotel costs: in £ Sterling (error bars represent 95% CI).

*Bimaxillary and single jaw surgery groups combined*

![Graph showing the sensitivity analysis of in-patient hotel costs for bimaxillary and single jaw surgery groups combined. The graph displays the cost in £ Sterling at different rates: Base (£150/night), Low (£100/night), and High (£200/night). Error bars represent 95% CI.](image)

**Bimaxillary surgery**

![Graph showing the sensitivity analysis of in-patient hotel costs for bimaxillary surgery. The graph displays the cost in £ Sterling at different rates: Base (£150/night), Low (£100/night), and High (£200/night). Error bars represent 95% CI.](image)
**Single jaw surgery**

![Graph showing costs in £ (95% CI) for single jaw surgery at different price levels: Base (£150/night), Low (£100/night), and High (£200/night).]
Table 27  A sensitivity analysis of operating theatre costs in response to alternative assumptions about costs incurred.

<table>
<thead>
<tr>
<th></th>
<th>Base estimate</th>
<th></th>
<th>High estimate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
<td>95% CI</td>
</tr>
<tr>
<td>Bimaxillary and single jaw combined</td>
<td>£1,519.49</td>
<td>£1,317.43 to £1,721.56</td>
<td>£2,483.22</td>
<td>£2,162.97 to £2,803.47</td>
</tr>
<tr>
<td>Bimaxillary surgery</td>
<td>£1,728.16</td>
<td>£1,540.70 to £1,915.62</td>
<td>£2,838.10</td>
<td>£2,571.82 to £3,104.39</td>
</tr>
<tr>
<td>Single jaw surgery</td>
<td>£997.83</td>
<td>£891.99 to £1,103.67</td>
<td>£1,596.83</td>
<td>£1,490.99 to £1,702.67</td>
</tr>
</tbody>
</table>
Figure 21 to 23  Sensitivity analysis of operating theatre costs: in £ Sterling (error bars represent 95% CI).

Bimaxillary and single jaw surgery groups combined

Bimaxillary surgery
CHAPTER 2 RESULTS

Single jaw

Cost in £ (95% CI)

N = 6  Base estimate  High estimate
Table 28  A sensitivity analysis of total resource use in response to alternative assumptions about costs incurred.

<table>
<thead>
<tr>
<th></th>
<th>Base estimate</th>
<th>Low estimate</th>
<th>High estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Bimaxillary and single jaw combined</strong></td>
<td>£3,320.95</td>
<td>£3,061.95 to £3,579.95</td>
<td>£3,147.14</td>
</tr>
<tr>
<td><strong>Bimaxillary</strong></td>
<td>£3,576.08</td>
<td>£3,326.63 to £3,825.53</td>
<td>£3,386.08</td>
</tr>
<tr>
<td><strong>Single jaw</strong></td>
<td>£2,683.12</td>
<td>£2,464.62 to £2,901.62</td>
<td>£2,549.79</td>
</tr>
</tbody>
</table>
Figures 24 to 26  Sensitivity analysis of changes in total resource use in response to alternative assumptions about cost: in £ Sterling.

Bimaxillary and single jaw surgery groups combined

Bimaxillary surgery
*Single jaw surgery*

![Graph showing total cost in £ (95% CI) for single jaw surgery. The graph displays three estimates: base, low, and high, with corresponding confidence intervals.]
Table 29  Mean non-discounted and discounted total resource use.

<table>
<thead>
<tr>
<th></th>
<th>Base estimate</th>
<th>Low estimate</th>
<th>High estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
</tr>
<tr>
<td>Bimax and single combined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-discounted</td>
<td>£3,320.95</td>
<td>£3,061.95 to £3,579.95</td>
<td>£3,147.14</td>
</tr>
<tr>
<td>Discounted</td>
<td>£3,181.59</td>
<td>£2,930.93 to £3,432.25</td>
<td>£3,015.02</td>
</tr>
<tr>
<td>Bimaxillary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-discounted</td>
<td>£3,576.08</td>
<td>£3,326.63 to £3,825.53</td>
<td>£3,386.08</td>
</tr>
<tr>
<td>Discounted</td>
<td>£3,431.44</td>
<td>£3,191.74 to £3,671.14</td>
<td>£3,249.09</td>
</tr>
<tr>
<td>Single jaw</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-discounted</td>
<td>£2,683.12</td>
<td>£2,464.62 to £2,901.62</td>
<td>£2,549.79</td>
</tr>
<tr>
<td>Discounted</td>
<td>£2,552.95</td>
<td>£2,362.78 to £2,743.13</td>
<td>£2,429.84</td>
</tr>
</tbody>
</table>
Figures 27 to 29 Mean non-discounted and discounted total resource use: in £ Sterling (error bars represent 95% CI).

**Bimaxillary and single jaw surgery groups combined**

![Graph showing mean non-discounted and discounted total resource use for bimaxillary and single jaw surgery groups combined.](image)

**Bimaxillary surgery**

![Graph showing mean non-discounted and discounted total resource use for bimaxillary surgery.](image)
Single jaw surgery
Table 30  Comparison of 95% confidence intervals for total costs as derived by conventional and bootstrap methods.

<table>
<thead>
<tr>
<th></th>
<th>Base estimate</th>
<th>Low estimate</th>
<th>High estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-discounted</td>
<td>Discounted</td>
<td>Non-discounted</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>£3,320.95</td>
<td>£3,181.59</td>
<td>£3,147.14</td>
</tr>
<tr>
<td>(conventional method)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td>£3,061.95 to £3,579.95</td>
<td>£2,930.93 to £3,432.25</td>
<td>£2,902.81 to £3,391.47</td>
</tr>
<tr>
<td>(conventional method)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td>£3,116.36 to £3,601.07</td>
<td>£2,981.80 to £3,447.32</td>
<td>£2,954.55 to £3,417.39</td>
</tr>
<tr>
<td>(bootstrap method)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2.5 Cost per QALY

Table 31 presents the cost per QALY (non-discounted and discounted) for bimaxillary and single jaw groups combined and separately and for all three methods: RS, SG and TTO. Values ranged from a low cost of £205.79 per QALY gained (single jaw group; RS method; low estimate; non-discounted costs and QALYs) to a high cost of £1,073.78 per QALY gained (bimaxillary group; SG method; high estimate; discounted costs and QALYs). Figures 30 to 32 illustrate these data graphically. Table 32 gives 95% CI (from bootstrapping) for cost per QALY values. Table 33 and Figures 33 and 34 show the sensitivity analysis of cost per QALY, based on the assumption that the end of treatment (T5) utility value is not retained throughout life.
Table 31 Mean cost per QALY.

<table>
<thead>
<tr>
<th></th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Groups combined</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-discounted</td>
<td>£236.87</td>
<td>£224.48</td>
<td>£368.14</td>
</tr>
<tr>
<td>and QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discounted costs</td>
<td>£354.30</td>
<td>£335.75</td>
<td>£550.46</td>
</tr>
<tr>
<td>and QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bimaxillary</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-discounted</td>
<td>£243.77</td>
<td>£230.82</td>
<td>£375.39</td>
</tr>
<tr>
<td>costs and QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discounted costs</td>
<td>£367.00</td>
<td>£347.50</td>
<td>£565.03</td>
</tr>
<tr>
<td>and QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Single jaw</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-discounted</td>
<td>£216.56</td>
<td>£205.79</td>
<td>£346.85</td>
</tr>
<tr>
<td>costs and QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discounted costs</td>
<td>£316.35</td>
<td>£301.10</td>
<td>£507.25</td>
</tr>
<tr>
<td>and QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 32  Mean cost per QALY with 95% confidence intervals from the bootstrap method (combined surgery group only).

<table>
<thead>
<tr>
<th></th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Non-discounted costs and QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (conventional analysis)</td>
<td>£236.87</td>
<td>£224.48</td>
<td>£368.14</td>
</tr>
<tr>
<td>95% CI (bootstrap)</td>
<td>£171.35 to £362.89</td>
<td>£162.64 to £343.76</td>
<td>£264.56 to £575.45</td>
</tr>
<tr>
<td>Discounted costs and QALYs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (conventional analysis)</td>
<td>£354.30</td>
<td>£335.75</td>
<td>£550.46</td>
</tr>
<tr>
<td>95% CI (bootstrap)</td>
<td>£257.17 to £547.38</td>
<td>£244.00 to £517.72</td>
<td>£397.02 to £861.40</td>
</tr>
</tbody>
</table>
**Figure 30**  Mean cost per QALY (non-discounted and discounted: base).

![Bar chart showing cost per QALY for different scenarios and groups.](chart1)

**Figure 31**  Mean cost per QALY (non-discounted and discounted: low).

![Bar chart showing cost per QALY for different scenarios and groups.](chart2)
Figure 32  Mean cost per QALY (non-discounted and discounted: high).
Table 33  Sensitivity analysis of cost per QALY gained: based on the assumption that the T5 utility value is not retained for life (groups combined).

<table>
<thead>
<tr>
<th>T5 utility value retained until (years of age)</th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Non-discounted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>£236.87</td>
<td>£224.48</td>
<td>£368.14</td>
</tr>
<tr>
<td>65</td>
<td>£293.37</td>
<td>£278.02</td>
<td>£455.95</td>
</tr>
<tr>
<td>55</td>
<td>£385.26</td>
<td>£365.10</td>
<td>£598.76</td>
</tr>
<tr>
<td>45</td>
<td>£560.97</td>
<td>£531.61</td>
<td>£871.85</td>
</tr>
<tr>
<td>Discounted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>£354.30</td>
<td>£335.75</td>
<td>£550.46</td>
</tr>
<tr>
<td>65</td>
<td>£403.76</td>
<td>£382.62</td>
<td>£627.30</td>
</tr>
<tr>
<td>55</td>
<td>£487.23</td>
<td>£461.72</td>
<td>£756.99</td>
</tr>
<tr>
<td>45</td>
<td>£651.97</td>
<td>£617.83</td>
<td>£1,012.94</td>
</tr>
</tbody>
</table>
Figure 33  Sensitivity analysis of cost/QALY gained: based on the assumption that the T5 value is not retained for life (non-discounted values): groups combined

Figure 34  Sensitivity analysis of cost/QALY gained: based on the assumption that the T5 value is not retained for life (discounted values): groups combined
4.2.6 Patients' comments

Comments at T1:
“I wouldn’t even have contemplated treatment had death been a consideration” (SG questions)

“Death does not bother me. I am a Buddhist and death is not something that we fear.” (Related to SG)

“Death doesn’t worry me - after all you wouldn’t know anything about it would you?” (Related to SG)

Comments at T2:
“Since I had braces fitted I have looked at other people’s teeth and realise that mine are not that bad really” (TTO value increased from 0.85 at T1 to 0.95 at T2)

Comments at T4:
“I’m still scared that the swelling is going to stay for ever, even though everyone tells me it will go. I could not live with that for the rest of my life.”

Comments at T5:
“I’m as near to perfect health now as I could hope to achieve. I would have to give the highest score. I wouldn’t trade anything at all for what I’ve got now.”

“I wouldn’t want perfect health - I’m happy as I am. If you were perfect you wouldn’t have anything to aim for...... and it’s good to still have something to complain about!”

“I’m not sure I did the right thing in having surgery now. I didn’t really have a serious problem” (patient whose pre- and post-treatment values were 0.95)
"I feel confused, on one hand I think I look good now and I am more confident, but on the other hand I just can’t get used to my face, it’s almost as if it doesn’t belong to me. It takes a lot of adapting to. Can you do this interview again in a years time and see if I have changed then?"

"I honestly didn’t realise how much more orthodontics I would need after the surgery, it has made me really depressed over the last few months and I think it will take a while to get over it. I sort of thought that surgery was the end point and it isn’t. I also didn’t realise how much the surgery would restrict what I can eat, even now when the braces are off."


5.0 WILLINGNESS-TO-PAY RESULTS

The experimental and control group subjects already described in the previous section also provided the data for the willingness-to-pay study. Table 34 presents test-retest data. Table 35 presents the WTP values and Table 36 shows a borderline significant difference (p=0.051) between the two groups. The correlations between WTP and utility values are shown in Table 37, with significant correlations between WTP and the SG (p=0.04) and TTO (p=0.02). The same table shows non-significant correlations between WTP and Townsend deprivation scores. No significant relationships between WTP and age were found (Table 38). WTP values also appeared quite stable with no significant changes between T1 and T5 for the longitudinal group (Table 39).

Table 34  Reliability test-retest data.

<table>
<thead>
<tr>
<th>Pearson correlation</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.82**</td>
<td>0.81</td>
</tr>
</tbody>
</table>

**p<0.01

Table 35  Willingness-to-pay values.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>95% CI</th>
<th>Range</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>£6,953.70</td>
<td>£5,529.17 to £8,378.23</td>
<td>£0-£15,000</td>
<td>£5,000</td>
</tr>
<tr>
<td>Control</td>
<td>£8,275.00</td>
<td>£7,218.53 to £9,331.47</td>
<td>£0 - £15,000</td>
<td>£8,000</td>
</tr>
</tbody>
</table>

NB: No significant gender differences
Table 36  Comparison of WTP between the two groups.

<table>
<thead>
<tr>
<th></th>
<th>U value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental vs Control</td>
<td>1767.0</td>
<td>0.051</td>
</tr>
</tbody>
</table>

U = Mann-Whitney U value

Table 37  Relationship between WTP and utility values and WTP and Townsend deprivation scores (experimental group only).

<table>
<thead>
<tr>
<th></th>
<th>Spearman correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>-0.30</td>
<td>0.04*</td>
</tr>
<tr>
<td>TTO</td>
<td>-0.34</td>
<td>0.02*</td>
</tr>
<tr>
<td>Townsend deprivation score</td>
<td>-0.16</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Table 38  Correlations between WTP values and age.

<table>
<thead>
<tr>
<th></th>
<th>Spearman correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>0.06</td>
<td>0.64</td>
</tr>
<tr>
<td>Control group</td>
<td>0.21</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Table 39  Change in WTP in the longitudinal group (T1/T5).

<table>
<thead>
<tr>
<th>Mean value at T1</th>
<th>Mean value at T5</th>
<th>Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(95% CI)</td>
<td>(95% CI)</td>
<td>(p-value)</td>
</tr>
<tr>
<td>£6,318.18 (£3,898.70 to £8,737.66)</td>
<td>£7,761.90 (£5,378.47 to £10,145.34)</td>
<td>-0.91 (0.37)</td>
</tr>
</tbody>
</table>

*Z value = Wilcoxon signed-rank test Z value*
6.0 DISCUSSION

6.1 GENERAL DISCUSSION

Three methods were used to determine utility values: the RS, SG and TTO. Overall, the methods were found to be acceptable to the respondents who appeared to understand the hypothetical situations and argued their utility choices with sound reasoning. Only one set of data for the control group could not be used as it was not complete. However, it must be borne in mind that the control group was a convenience sample.

It must also be borne in mind when interpreting the data for this study that the patient population from the Eastman Dental Institute cannot be assumed to be the same as those in any other centre. This is particularly relevant in that the unit is a tertiary referral centre and there are certain experts available who are not available in all units, in particular clinical psychologists and a liaison psychiatrist. The data, therefore, relates only to patients from this Institute and it would be necessary to replicate the study in other units to determine whether similar values are given.

The SG has always had support because it is based on the von Neumann and Morgenstern theory (von Neumann and Morgenstern, 1953). However, there are problems when using this method to investigate chronic conditions because the treatment for most chronic conditions does not approximate the gamble. Often there is no product that will make the patient perfectly healthy nor is there a treatment which is likely to kill them, therefore, the gamble is dissimilar to the real choices available to the individual (Kaplan, 1998). These issues are particularly pertinent for orthognathic patients and the chronic condition of dentofacial deformity. However, a conventional "perfect health/death" option was recommended for this study (Torrance, Personal communication). The reasoning for this is grounded in two main theories:
• The values could then be compared with those from other medical interventions. If the anchors were not "perfect health" and "death", the utility values would have limited use because they could not be compared with other studies which have used this method.

• The oral cavity is part of overall health and should be considered in the overall concept of an individual's health. Traditionally there has been a tendency to treat it as a landmark which happens to be located within the body and, as such, the oral cavity has been seen as separate to the body and the individual. This approach is now being questioned and this has given rise to new concepts in which oral conditions are linked to diseases in other body locations and to health outcomes of quality of life (Locker, 1997). Locker (1997) challenged the distinction often made between general and oral health and introduced the issue that our focus should be, not on the oral cavity itself, but on the individual and the way in which the oral condition affects health and well-being.

If orthognathic treatment is to be seen on a par with other medical interventions, it is important that the dental profession uses the same techniques and does not perceive it as a "dental" problem. It is also important to acknowledge the major impact that dentofacial deformity can have on quality of life. It is not an isolated incident for patients to consider suicide as a result of their appearance and associated problems. In addition, all patients accept a risk of death (albeit a small risk) in agreeing to undergo surgery under general anaesthetic.

It is possible to alter the anchor state (death) and the values derived can then be "chained" back to the perfect health/death scale, provided the reference state is valued against perfect health/death in another gamble. It is, however, difficult to choose an alternative anchor state which would be perceived in the same way by all respondents. If the "worst facial and/or dental problems imaginable" was selected as an anchor, this would vary between individuals. This then has major implications if the anchor state is not consistent. Until some method of grading facial appearance is developed (perhaps...
in a similar manner to the IOTN aesthetic component), this is a particularly difficult issue.

There are a number of other studies which have used the conventional SG approach for conditions which are chronic and not life threatening. For example, Wells and Sherbourne (1999) used the TTO and SG to assess utility values for a wide range of chronic medical conditions. Guyatt et al. (1999) used the SG technique in their study of patients with chronic airflow limitation and found a baseline mean value of 0.55 which did not change significantly following rehabilitation treatment. Drummond et al. (1987) described the SG method in detail using a perfect health/death scale and the health states which were described included two non-threatening conditions: blindness and mental retardation. Goossens et al. (1999) used the RS and SG for patients suffering from fibromyalgia and chronic back pain. However, they did conclude that the SG questions were hypothetical with that group of patients and that it was a far from realistic situation. Their mean and median baseline SG scores were 0.95 which is higher than in this study. This may suggest that chronic pain is an easier situation to live with than disfigurement but it must also be considered that the age group in the current study is younger, such that patients would have to cope with the problem for longer and this may influence utility values.

It remains a matter for debate as to whether the SG is an appropriate method for this group of patients and a number of subjects did comment that the situation was unrealistic. It seems likely that the TTO provided more realistic responses and respondents appeared to experience less difficulty with this technique. In future, consideration should be given to using the TTO rather than the SG with orthognathic patients. One of the main problems associated with the TTO, however, is that the technique assumes that a subjects’ perception of time is linear and this may not be true (Boyle and Torrance, 1984). Stiggelbout et al. (1995) also questioned whether the TTO was understood by the average patient in view of their study which found that many oncology patients were not willing to trade at all. There was, however, no
evidence in the current study to suggest that the respondents did not understand the questions being asked.

6.2 COMPARISON OF PRE-TREATMENT UTILITY VALUES FROM BOTH CONTROL AND EXPERIMENTAL GROUPS

Table 9 shows that the gender distribution was approximately two-thirds female and one-third male for both the experimental and control groups. There was no significant difference in the age distribution for the two groups. In some utility studies (Dolan et al., 1996), values were found to differ by age and gender. Therefore, the control group was selected to be comparable with the experimental group.

6.2.1 Test-Retest

Table 10 illustrates the test-retest results. The ICC values were considered acceptable for all three methods (varying between 0.64 and 0.74). Similar levels of repeatability were found for all three methods using the Bland and Altman method (standard deviation of the differences were RS 14.4; SG 0.10; TTO 0.13). These values are in keeping with those quoted in previous papers and support the concept that individual utility values are relatively imprecise (Torrance, 1976; Torrance et al., 1982).

6.2.2 Comparison of mean values

Table 11 shows the mean and median utility values and the 95% confidence intervals for both groups. The mean utility value for the rating scale was 63.49 for the experimental group and 55.38 for the control group. Mean values for the standard gamble were 0.76 for the experimental and 0.83 for the control group. Mean values
for the time trade-off were 0.68 for the experimental group and 0.74 for the control group. There were no significant differences between the mean values for the two groups for either the SG or the TTO. However, there was a significant difference for the RS (p=0.02) with the mean value for the control group being significantly lower than that for the experimental group (Table 12). Nevertheless, it is important to bear in mind that the sample size for the experimental group was lower than ideal for the SG and TTO methods according to the sample size estimation. These results are another example of the rating scale behaving in a different manner to the SG and TTO and reinforces the need to include either the SG or TTO with the RS. Another factor which must also be borne in mind is that the control group was a convenience sample and, as such, there may be some bias. However, the difficulties of recruiting a representative sample of the population in a study of this type are yet to be resolved.

The three methods produced different values, as previously noted by Mooney (1994). Perhaps not surprisingly, the highest utility value was provided by the standard gamble method and the lowest by the rating scale. The lower value from the rating scale reflects findings from previous research (Torrance, 1976). It seems probable that TTO values were lower than those from the SG because, although most respondents were quite willing to consider losing a number of years of their life when they had 50 years to consider, the extent of life/death gamble they were prepared to accept was, not surprisingly, much more limited.

Also, of note is the difference between the standard deviations for the two groups. The standard deviation for the control group was lower than the experimental group for all three methods but was markedly lower for the SG and TTO. This contradicts findings by Torrance (1976) who found standard deviations which were higher for the general public. It must be noted, however, that in this study the range of values for the experimental group was noticeably wider than that for the control group. This may show that the control group are giving “cultural norms” regarding facial aesthetics and that there is less individual variation than in the experimental group.
There was some concern that utility values may be influenced by the age of the patient in view of the relatively wide range (16 to 42 years). However, Table 13 shows that the correlations between age and utilities were low and not of a significant level. Only the RS for the experimental group came close to showing a significant correlation. This was a positive correlation, i.e. as age increases, so does the RS value. However, on the whole, these findings offer support to the argument that utility values are true measures of strength of preference.

6.2.3 Method agreement

Table 14 shows that for both the experimental and control groups there were marked differences between all three methods. This supports previous research which found different utility values using different methods (Mooney, 1994). The coefficient of reproducibility was lower for the SG/TTO than for the SG/RS and TTO/RS. This suggests that the SG and TTO have better method agreement than either method has with the RS. Therefore, utility values derived by different methods are not interchangeable and this illustrates the importance of quoting the method used as well as the utility value obtained. The lack of comparability of the rating scale with other methods has also been noted in previous studies. Torrance (1976) believed that this may be due to the fact that no gamble is involved and suggested that the rating scale should always be combined with one of the other methods and not used in isolation. In a later paper, Torrance (1987) proposed that the rating scale may be related to the other methods by means of a power curve rather than a linear function.
6.2.4 General discussion

Interpretation of results obtained in utility studies is always difficult. Information is achieved primarily by comparing the values with those obtained for other health states. For example, the following utility values were obtained using a conventional TTO method by Torrance (1987):

- Some physical and role limitation with occasional pain 0.67
- Hospital dialysis 0.56
- Anxious and lonely most of the time 0.45
- Blind or deaf or dumb 0.39
- Hospital confinement 0.33

Comparing these results with the present study, where the mean utility value using the TTO was 0.68 for the experimental group, suggests that dentofacial deformity rates on a par with physical and role limitation but is not considered as severe a problem as the other health states in this list.

A study by Downer et al. (1997) using the SG approach to assess different stages of oral cancer found utility values of:

- Precancer (assuming full recovery) 0.92
- Stage 1 oral cancer (assuming full recovery but regular checks) 0.88
- Stage 2 oral cancer (assuming the cancer is life threatening) 0.68

Again, when the results of the current study (SG value of 0.76 for the experimental group and 0.83 for the control group) are compared with these values, dentofacial deformity is considered a worse health state than either of the lesions when full recovery is assumed but not as great a problem as a life threatening lesion. Obviously, one issue which must be borne in mind in this situation is that oral cancer tends to occur in an older age group, which may influence the results.

One of the underlying aspects to this study was that it was felt there may be a problem ensuring that members of the public were adequately informed when determining utility values for dentofacial deformity. Unlike other medical interventions, such as treatment for arthritis or angina, this is an area with which the majority of lay people
may not be familiar. The fact that there were no significant differences between the two groups for two of the three methods suggests that members of the public understood the health state descriptions provided and that general public utility values may be considered a satisfactory source of data.

6.3 LONGITUDINAL UTILITY ANALYSIS

6.3.1 General discussion

Figures 8 to 10 and Table 16 show the mean utility values throughout treatment for the 21 subjects in the longitudinal group. Examples of individual plots are shown in Appendix 2 and the great variability between subjects is of note. This is in keeping with suggestions by other authors that single measurements are not highly precise (Torrance, 1976; Sackett and Torrance, 1978; Torrance et al., 1982). In addition, the marked variation within individual subjects but between the three methods is evident, this also supports the findings of previous research.

It is evident from Figure 8 that the RS produces a different pattern of values as compared with the SG and TTO. The RS starts with lower utility values which then increase at T2 and further at T3. In contrast, the SG and TTO show decreases at T2 and T3 when compared with T1. This provides further evidence that although the SG and TTO are thought to produce true utility values, the RS may not. It is unlikely that the health state was actually improved at T2 and T3, however, it is probable that the RS is unable to differentiate between the health state being valued at that point in time and the anticipation of a better health state as a result of treatment. In contrast, the SG and TTO were able to measure the true health state at that point and any potential improvements in the future were excluded from consideration. Both the SG and TTO showed a small increase in values between T3 and T4 although the RS remained approximately the same. It is not surprising that utility values were still low at this
point because surgery was a relatively recent event. All three methods showed an increase from T4 to T5.

There were no significant differences between utility values for the bimaxillary and single jaw groups at T1 or T5 (Table 17). All three methods showed a significant increase in values between T1 and T5 for the bimaxillary and single jaw groups combined and for the bimaxillary group (Table 18). Interestingly, the single jaw group did not show any significant change between T1 and T5. However, these results must be treated with some caution because the single jaw surgery group was very small in number (n=6) and, in addition, there were two individuals in that group who were uncertain about the results of their treatment. Indeed, one respondent gave exactly the same utility values throughout and another gave a lower utility value at T5 as compared with T1 (resulting in negative QALYs). Although this value may increase as the patient becomes used to their new appearance, this cannot be assumed. Therefore, until proven otherwise, it must be assumed that the value is permanent. It is intended, however, to re-interview all subjects again at 1 year to ascertain whether this is the case. All subsequent results were presented for the groups combined and separately, although acknowledging that the single jaw group data must be treated with caution.

6.3.2 Calculation of QALYs

Tables 19 to 21 show the QALYs lost or gained during treatment and Table 22 shows the total QALYs gained as a result of treatment. It is not unreasonable to assume that there will be a reduction in utility values during treatment (so-called “disutility”), whether this is associated with orthodontics, surgery or both. Table 19 shows the values for bimaxillary and single jaw groups combined. For the rating scale method, QALYs were gained during treatment (mean value of 0.27 QALYs) whereas the standard gamble and time trade-off measured negative QALYs during treatment (-0.10 QALYs for both). Clinical judgement would suggest that QALYs are likely to be lost and this further reinforces the difference between the RS and the SG and TTO. It
appears that with the RS, respondents were unable to divorce the concepts of their current health state from the improvements which they were anticipating.

The QALYs lost or gained during treatment were largely unaffected by discounting due to the short time periods involved. However, discounting does have a marked effect on the QALYs gained as a result of treatment. For example in Table 19, using the TTO method, non-discounted and discounted QALYs gained as a result of treatment were 9.03 and 5.77 respectively. This is due to the fact that discounting occurs over a period of around 50 years for this calculation.

Total QALYs gained as a result of treatment varied for the three methods (Table 22). For the bimaxillary and single jaw surgery groups combined, the RS showed the greatest number of QALYs gained (14.02 non-discounted and 8.98 discounted) with the SG and TTO providing similar values (8.13 and 5.12 for SG; 8.93 and 5.67 for TTO). The single jaw group showed lower values than the bimaxillary group for the RS and TTO, although not the SG. As discussed previously, these findings must be treated with some caution, although it is an interesting question as to whether patients undergoing single jaw surgery do experience less gain than those undergoing bimaxillary surgery and this should be looked at in more detail in the future. The greater QALYs gained in the bimaxillary group may be a reflection of the fact that these patients presented with more severe malocclusions initially and therefore gained more improvement due to surgery. However, there was no significant difference between the two groups with respect to presenting utility values at T1 (Table 17).

Overall, there was a positive change as a result of treatment, with the QALYs gained far outweighing any negative values. Figures 11 to 13 represent the QALYs gained/lost as measured by the TTO method and reinforce the small impact of QALYs lost during treatment as compared with the overall gains. It is important, however, that clinicians acknowledge this issue of “disutility” during treatment and explain the negative aspects to patients.
95% CI for total QALYs gained were compared using the conventional and bootstrap methods (Table 23). For all three methods (RS, SG and TTO), the bootstrapped values showed 95% confidence intervals which were slightly narrower than those calculated conventionally. However, even with a small sample size, the values were remarkably similar using both methods, as postulated by Thompson and Barber (2000). It is, therefore, reassuring that the sample is providing data which can be utilised reliably.

Calculation of QALYs gained as a result of treatment does make the assumption that the end of treatment utility value is retained over the rest of the patient’s life and this may not be true. It may decrease as the patient becomes older and facial/dental aesthetics becomes less important to them. This could form the basis for future study to establish what period of time the end of treatment utility value is maintained over. Figure 14 and Table 24 present the results of a sensitivity analysis assuming that the T5 utility value is not retained for life but returns to pre-treatment values at 45, 55 and 65 years respectively. It would also be feasible to undertake sensitivity analyses of the QALYs lost/gained during treatment with different assumptions regarding the pattern of changes. However, the values were so small compared with the overall gains due to treatment that it was felt there was minimal benefit in doing so. In future studies it would be interesting to include an additional interview immediately following surgery to determine the pattern of changes in the T3/T4 time period. The logistics of this would have proved difficult in this study as the in-patient beds are away from the main dental hospital site. In addition, there are ethical considerations as some patients may find this intrusive in the immediate post-operative phase.

This study is one of very few which has studied changes in utility values during a course of treatment. The majority of studies report values only before and after intervention. Utility values and QALYs gained as a result of treatment have now been reported for a number of different conditions and interventions. For example, Cheng and Niparko (1999), in their meta-analysis of cochlear implants, reported utility gains of between 0.05 and 0.41. In a study looking at anterior cruciate ligament repair in
young adults, surgery resulted in 5.10 QALYs gained in the first seven years after injury. However, it must be borne in mind that these values were determined from a group of students and not from patients (Gottlob et al., 1999). Heudebert et al. (1997) utilised Markov modelling to compare two strategies for managing oesophagitis: laparoscopic Nissen fundoplication (LNF) and omeprazole. Over a time horizon of 5 years they found 4.33 QALYs gained for both treatment methods.

6.3.3 Calculation of costs

Appendix 3 shows the resource use for orthognathic procedures based on costs at the Eastman Dental Hospital and UCLH. Resource use was calculated as accurately as possible for each individual patient. This was a time consuming task and was possible due to the relatively small sample size. Larger studies may preclude exact individual calculations. It must also be borne in mind that all the patients in the study were treated by junior orthodontic staff (approximately half by registrars and half by senior registrars) and this may have had an effect on orthodontic resource use. In the majority of orthodontic units outside dental schools it will be predominantly consultants and senior trainees who treat orthognathic patients. However, the surgical and in-patient costs represent the greatest proportion of the total costs. This included in-patient costs, theatre time, surgeon costs and consumables (including titanium mini-plates/screws).

In view of the fact that the treatment for each respondent was calculated individually, complications which resulted in additional resource use could also be included. Those which had the greatest impact on resource use were related to theatre time. For example, one patient required re-operation and then a subsequent day unit procedure for wound exploration; two patients required removal of plates/screws and one required closure of a palatal fistula created following a midpalatal split. Other sources of additional resource use were accrued through an additional rhinoplasty procedure (one patient) and placement of dental implants (two patients).
The perspective of this study was that of a third-party payer (the NHS in this instance) and indirect costs (for example, costs to the patient due to time off work) were not included. This may have influenced the results although it was considered that because many of the respondents were students this would have less impact than if they were in employment and suffered financially due to time away from their employment.

Sensitivity analyses are required in a study of this type to allow for the fact that it is never possible to estimate costs with total accuracy (Tables 25 to 28). Table 25 and Figures 15 to 17 present the sensitivity analysis of orthodontic resource use. The base costs represented the individual “bottom up” calculations as shown in Appendix 3. The other estimates involved using the £28.00 per visit cost provided by the Eastman Dental Hospital (UCLH Trust) and the figure of £42.00 per visit which represents the UK costing for dental out-patient visits. For the two surgery groups combined, the mean value could vary from a base estimate of £707.99 (95% CI £682.11 to £733.87) to a high estimate of £1,262.00 (95% CI £1,169.03 to £1,354.97). The values for the single jaw surgery group were in fact slightly higher than for the bimaxillary surgery group although this was not significant. Similarly there was no significant difference between the orthodontic costs for those patients treated by registrars as compared with those treated by senior registrars.

The sensitivity analysis of in-patient hotel costs used a base cost of £150/night with costs of £100/night and £200/night representing the low and high estimates respectively (Table 26 and Figures 18 to 20). For the two surgery groups combined, the estimates ranged from a low estimate of £347.62 (95% CI £303.36 to £389.88) to a high estimate of £695.24 (95% CI £610.72 to £779.76). Perhaps not surprisingly, the single jaw surgery group showed lower values than the bimaxillary group (mean base costs of £400.00 and £570.00 respectively). The surgery involved for the single jaw group may be associated with less morbidity and therefore the time spent as an in-patient is likely to be less.
The sensitivity analysis of operating theatre costs used a base cost derived by "bottom up" costing and the high estimate used the sessional cost supplied by UCLH Trust (Table 27 and Figures 21 to 23). There was no appropriate low value to use in this analysis. Again, the costs for single jaw procedures were lower than for bimaxillary surgery and this reflected the difference in theatre time and staff costs required for the two procedures. For the bimaxillary group, the resource use varied from a base estimate of £1,728.16 (95% CI £1,540.70 to £1,915.62) to a high cost of £2,838.10 (95% CI £2,571.82 to £3,104.39) and for the single jaw group from £997.83 (95% CI £891.99 to £1,103.67) to £1,596.83 (95% CI £1,490.99 to £1,702.67).

In view of the findings for in-patient hotel costs and theatre costs, it is not surprising that the single jaw group gave lower values than the bimaxillary group for total resource use (Table 28 and Figures 24 to 26). For the two groups combined, resource use varied between a low estimate of £3,147.14 (95% CI £2,903.09 to £3,391.19) to a high estimate of £5,161.34 (95% CI £4,726.38 to £5,596.30). The mean values for the single jaw group varied between £2,549.79 to £4,297.42 and for the bimaxillary group between £3,386.08 and £5,506.90. These values came as some surprise as it had been assumed that total resource use for orthognathic treatment would be considerably higher than this. Table 29 and Figures 27 to 29 present non-discounted and discounted total resource use and Table 30 shows the 95% confidence intervals as derived from conventional and bootstrap methods. Reassuringly, the values were very similar for the two methods.

It is difficult to find similar interventions with which to compare the present study. In addition, it is often comparative costs for two possible interventions which are reported. In their study comparing hysterectomy and endometrial resection in women with menorrhagia, Sculpher et al. (1993) reported mean costs of £560.05 for endometrial resection and £1,059.73 for the hysterectomy group. Although these values are lower than the costs for this study, it must be considered that orthognathic treatment involves resource use over a 2 year period rather than a one-off in-patient stay. In addition, the orthognathic surgical procedures (particularly bimaxillary
surgery) require more theatre time than the surgical procedures in question in the study by Sculpher et al. (1993, 1996). Gottlob et al. (1999) found that the total cost for anterior cruciate ligament reconstruction was US$11,768. However, different approaches to medical care between the United States and the United Kingdom must be considered. In their study of cochlear implants, Palmer et al. (1999) reported mean costs of Can$37,405, although approximately 60% of this cost was the implant device itself. In the study by Heudebert et al. (1997), comparing two strategies for managing severe oesophagitis, costs were noted to be higher for the surgical intervention at US$9,482 versus long-term non-surgical treatment at US$6,053.

6.3.4 Cost per QALY

Costs per QALY vary enormously, from interventions which actually save money to those interventions which cost hundreds of thousands of pounds for each QALY gained. The mean cost per QALY varies depending on the sensitivity analysis and also whether or not values are discounted. In order to determine whether the cost to benefit ratio is acceptable, it is generally considered that interventions which produce a cost per QALY below £10,000 to £25,000 per QALY gained should be funded (Sculpher, Personal communication).

Costs per QALY are presented in Table 31 with 95% confidence intervals in Table 32. These results indicate there can be little doubt that orthognathic treatment comes into the category of interventions which are suitable for funding. The cost per QALY was still much below these levels even if the T5 utility value is not retained throughout life, as shown in Table 33 and Figures 33 and 34.

The maximum cost per QALY calculated was £1,073.78 (for the bimaxillary group; SG method; high estimate; discounted costs and QALYs). The SG tended to give fairly high costs per QALY by virtue of the fact that this method also gave lower QALYs gained (Figures 30 to 32). This makes orthognathic treatment an intervention
which provides a good outcome at relatively low cost. This can be explained, in part, due to the costs being a “one-off” occurrence, with patients not requiring additional expensive treatments after the main treatment process is completed. In addition, the patients are relatively young, therefore the QALYs gained are high because the future life expectancy is generally around 50 years. Whilst acknowledging that QALY league tables should be treated with caution, it is reassuring to know that orthognathic treatment should fare well.

Again, finding studies which confer similar benefits as those from orthognathic treatment is difficult. A number of studies looking at cochlear implants were felt to be comparable in that the patients’ quality of life is enhanced rather than the procedure being life saving. For example, in their meta-analysis of the cost-utility of cochlear implants in adults, Cheng and Niparko (1999) found a cost per QALY of US$12,787. These figures were felt to compare favourably with interventions which are funded by third-party payers in the United States at the present time. A similar study of cochlear implants by Carter and Hailey (1999) found costs per QALY of Aus$5,070-$11,100 for children and Aus$11,790-$38,150 for profoundly deaf adults. One of the features which raises the cost of this intervention relative to orthognathic treatment is the high cost of the cochlear implant itself.

Another study which looked at a chronic non-life threatening condition was that by Ebell et al. (1997) who investigated the management of dyspepsia. Seven treatment strategies were compared and three were similarly cost-effective at a cost per QALY ranging from US$1,198 to US$1,288, although one strategy was noted to be associated with more recurrences than the other two. It was, therefore, concluded that two of the strategies could be recommended for funding.

Often two treatment methods are being compared and it is then the incremental cost per QALY that is quoted, i.e. the additional amount to treat by one method over the other. Financiers must then decide whether the additional costs are justified to fund a certain treatment. Gottlob et al. (1999) compared operative versus non-operative
management for acute anterior cruciate ligament repairs and found a marginal cost effectiveness of US$5,857 per QALY. Sculpher (1998) showed that treating menorrhagia by abdominal hysterectomy rather than by endometrial resection was more effective if purchasers were willing to pay an additional cost of £6,500 per extra QALY gained. Porter et al. (1997) found an incremental cost effectiveness of Can$7,100 for surgery versus stereotactic radiosurgery in the treatment of cerebral arteriovenous malformations.

In contrast to interventions considered worth financing, Forbes et al. (1999) found that the benefits of interferon beta-1b in treating multiple sclerosis were very low relative to its cost and estimated that in order to treat sufficient patients to prevent one individual becoming wheelchair bound would cost over one million pounds. They therefore concluded that money would be better spent on other ways of improving quality of life rather than funding treatment of this type. Other studies have produced findings which leave funding of the intervention open to some doubt. For example, Leal et al. (1999) looked at the cost utility of screening intravenous drug users for hepatitis C and found a cost per QALY of £9,300. However, the sensitivity analysis gave such a wide range of values that they concluded that there was insufficient evidence to implement the recommendations.

Cost per QALY figures are, at best, estimates of averages and they are intended to be used as a guide for resource allocation. It is important to acknowledge that QALYs and costs per QALY vary greatly depending on methods used, the time and place of the survey and changing techniques. This technique does, however, show promise for use in the dental field.
6.4 WILLINGNESS-TO-PAY

The willingness-to-pay method has not been used frequently in the field of dentistry. However, respondents answered the questions readily and there were no refusals to complete this section of the interview. This study therefore confirmed that the willingness-to-pay method was acceptable to both control group respondents and patients requesting treatment for dentofacial deformity. There was a mean WTP value of £6,953.70 for the experimental and £8,275.00 for the control group (Table 35) with a difference which was of borderline significance ($p=0.051$). It may be that the control group provided a higher value because they knew there was no possibility of them having to consider paying for this form of treatment, whereas the experimental group may have given a more considered, realistic amount because they appreciated that payment for treatment may have been necessary in certain situations. Interestingly, several respondents likened their willingness-to-pay values with how much they would be prepared to incur as a student loan and others compared the value with how much they would be prepared to pay for a new car.

Comparison with other studies is very difficult. There are still a limited number of WTP studies in health care and many of these have been undertaken in the United States or in other countries where payment for medical care is the norm. This makes the concept of WTP easier for respondents to understand. In addition, some studies have estimated WTP as a percentage of household income, a technique which could not be utilised in this study. For example, Thompson (1986) noted that individuals were willing to pay up to 22% of their household income for a cure for rheumatoid arthritis.

The correlations between WTP and utility values for the experimental group (Table 37) were significant and in the expected direction (-0.30 $p=0.04$ for SG; -0.34 $p=0.02$ for TTO). This means that those individuals who gave the lowest utility values (i.e. were prepared to risk most for treatment) were also willing to pay the most. This suggests that willingness to pay can be used as a measure of strength of preference. O’Brien and Viramontes (1994) found a similar correlation between the SG and WTP.
in their study of patients with chronic lung disease. Correlations were not determined for WTP and the RS as it was felt that the RS does not involve a measure of strength of preference in the same way as the SG and TTO.

Although the income for each household was not recorded in this study, a measure of socio-economic background was estimated using postcodes to obtain a Townsend deprivation score (Townsend et al., 1986). The relationship between Townsend deprivation score and WTP was in the expected direction (i.e. as social deprivation increased the WTP was less) but this was not significant (p=0.28). In this study, the WTP was therefore not significantly influenced by ability to pay, which again supports the proposal that WTP can be used as a measure of strength of preference (Donaldson et al., 1995a). It is recognised that the Townsend index is not without flaws because there is no individual measure; it gives only an area measure of residency and does not pertain to the individual. However, this was the best indicator available in the absence of household income. In view of the fact that postcodes were not obtained for the control group, Townsend scores could not be calculated and correlations for this group are therefore not presented.

When WTP values were compared at T1 and T5 for the longitudinal group there was no significant difference between the two values (Table 39). This suggests that the respondents were providing consistent responses.

The use of WTP remains a controversial issue. However, this study suggests that it may be usefully combined with utility analysis when looking at strength of preference for health care intervention. Larger sample sizes would be useful for further investigation. It would also be useful to compare different forms of health care intervention in the same study. For example, comparison could be made of the willingness-to-pay for orthognathic treatment and adult orthodontic treatment.
7.0 CONCLUSIONS

Utility analysis

- The results of this study suggest that valid utility values can be obtained from members of the general public, even when they are not familiar with the health state in question, provided it is described adequately. The method of using photographs to describe a health state is a novel approach which may be useful in future studies.

- It is recommended that for studies in the field of orthodontics and orthognathic surgery, the time trade-off and rating scale be used in preference to the standard gamble. There remains some doubt as to whether appropriate utility values were obtained using the standard gamble.

- For the bimaxillary and single jaw surgery groups combined, there was a significant increase in utility values following treatment, for all three methods.

- The mean non-discounted total QALYs gained as a result of treatment were 14.02, 8.13 and 8.93 for the RS, SG and TTO respectively (for the bimaxillary and single jaw groups combined).

- The mean cost per QALY varied between a minimum value of £205.79 and a maximum value of £1,073.78. This suggests an intervention which provides good benefits for a relatively low cost. The recommendation would be that orthognathic treatment is an intervention which should be funded.

- The bootstrap method confirmed the findings of conventional statistical analysis. This was reassuring with the small sample size involved.

- The cost-utility technique is one which should be explored further in the field of dentistry.
Willingness-to-pay

- There were significant correlations between WTP and the SG and TTO utility values. This suggests that WTP can be used to measure strength of preference.

- There was no correlation between WTP and Townsend deprivation score which indicates that WTP is independent of socio-economic background. This suggests that WTP may be independent of ability to pay.

- The WTP technique shows promise for future use in dental research.
8.0 SUGGESTIONS FOR FURTHER STUDY

The use of economic evaluation in dentistry is likely to increase in the future with the increased accountability to the providers of funds, whether this is in the NHS or the private sector.

It is intended to continue this research by undertaking a larger scale cost-utility analysis of orthognathic treatment, involving other units to increase the number of respondents. This will also allow differences between bimaxillary and single jaw surgery groups to be studied and reported with greater confidence.

It is also intended to carry out a similar study for combined adult orthodontics/restorative treatment, particularly those forms of treatment involving expensive restorative work, such as dental implants.
CHAPTER 3

QUALITY OF LIFE IN ORTHOGNATHIC PATIENTS

“Quality of life is defined as an individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person’s physical health, psychological state, level of independence, social relationships, and their relationships to salient features of their environment.”

(WHOQOL Group, 1996)
1.0 REVIEW OF THE LITERATURE

1.1 INTRODUCTION

"is it so surprising that patients with disabilities can find quality of life in the midst of their limitations, not only in spite of their limitations but, dare I say, sometimes even because of their limitations"

(de Lateur, 1997)

Quality of life (QOL) became an important issue in the United States in the 1970s when well-being in the general population was investigated (Brown and Gordon, 1999) and the term quality of life became a key word in the Medical Subject Headings of the US National Library of Medicine MEDLINE Computer Search System as recently as 1977. Since then interest in this field has increased enormously with over 1000 new articles indexed each year under the quality of life heading (Muldoon et al., 1998; Wood-Dauphinee, 1999).

QOL may be defined as “a person’s sense of well-being that stems from satisfaction or dissatisfaction with the areas of life that are important to them.” The concept is a difficult one because it means different things to different people and can also mean different things to the same person over a period of time (Sprangers and Schwartz, 1999). Health plays a large part in quality of life and this has become known as health-related quality of life (HRQL). HRQL has become increasingly important as researchers realised that traditional outcome measures are of little interest to the patient and that some form of “real life” outcome measure is required in the current health climate (Dijkers, 1999). To fully evaluate any healthcare intervention requires outcome measures of importance to the patient as well as to the clinician. Focusing concern from QOL to HRQL does have some limitations, especially with the acceptance that there is more to life than health and that individuals make decisions regarding health care not only on the basis of symptoms but also on external influences.
such as concerns of loved ones (Lawton, 1999). However, in most instances it is HRQL which is important in health care research, whilst acknowledging that it is not the only type of QOL (Spilker, 1996).

An underlying assumption of HRQL research is that, in addition to relieving clinical symptoms and prolonging survival, a primary objective of any intervention is the enhancement of quality of life (Berzon, 1998). As recently as 1987 it was noted that “few surgical trials consider quality of life variables as outcome measures.... Unless quality of life effects are quantified and reported in trials, they will be ignored or undervalued in health policy decisions” (O'Young and McPeek, 1987). The authors looked at six well known and respected general surgery journals for the year beginning 1981 and, of 99 therapeutic trials, 97 per cent made no mention of quality of life. They cautioned that work was required to develop appropriately designed quality of life studies and the last decade has seen major advances in this area.

Research into HRQL during the 1970s focused largely on construction and development of instruments. This continued during the 1980s and 1990s with the Short Form-36 (SF-36) questionnaire first appearing in 1992 and subsequently becoming probably the most widely used health status measure in the world. An increased interest in methodological issues also occurred during the 1980s and 90s with emphasis being placed on good psychometric properties. There has also been demand to reduce the length of quality of life instruments in order to reduce the time required for completion and improved statistical methods have done much to aid this (Wood-Dauphinee, 1999).

Despite major research efforts to date in the field of HRQL, there remains work for the future. There still exists no conceptual model or theory as a foundation for the HRQL construct and to explain the relationships between components. Improvements in this area would make it easier for researchers to select the appropriate instrument for their population group. The future is also likely to see more condition-specific measures and individualized measures such as the Schedule for the Evaluation of
Individual Quality of Life or SEIQoL (O’Boyle et al., 1992) and the Patient Generated Index (Ruta et al., 1994). It may even be feasible to incorporate HRQL measures into hospital databases.

1.2 QUALITY OF LIFE MODELS

Despite the increased interest in quality of life, there remains disagreement on exactly what it means and how it should be measured. This is even further complicated by some individuals who believe that quality of life is not quantifiable and must be assessed through qualitative study. Despite the lack of a definition, what is agreed is that HRQL should include those areas of concern to individual patients (Brown and Gordon, 1999; Jenkinson et al., 1999). HRQL is assumed to encompass many elements of an individual’s life which are not accessible to the doctor and it may, therefore, be argued that the patient is the best person to judge their own quality of life (Hunt and McKenna, 1992).

It is now generally, although not universally, accepted that HRQL includes a number of domains. Spilker (1996) proposed the following domains:

- physical status and functional ability;
- psychological status and well-being;
- social interactions;
- economic and/or vocational status and factors;
- religious and/or spiritual status.

Others have suggested a smaller numbers of domains. For example, Schipper et al. (1996) and de Graeff et al. (1999) quoted four domains: physical and occupational function; psychological state; social interaction and somatic sensation (i.e. disease symptoms).
Although it is generally agreed that HRQL is a multidimensional concept, there are two schools of thought. One group argues that measurement of HRQL is multidimensional such that a quality of life profile is produced; the argument is that measurements are only possible within a domain and not between domains. The SF-36 (Ware and Sherbourne, 1992; McHorney et al., 1993) and the Nottingham Health Profile (McEwen and McKenna, 1996) are examples of this. The other argument is that since people have to weigh up the very diverse attributes of health, it should be possible to elicit a single index value. It is within this second school that the EuroQol is grounded (EuroQol Group, 1990).

When new instruments are being developed, the researchers must decide whether each domain should be treated as a separate score, as in the SF-36, or whether the scores should be summed to produce a single score (Dijkers, 1999). If the latter approach is chosen, it must also be decided whether weighting is required or whether all domains are given equal weight. Weighting is itself open to criticism and certainly no approach is without fault (Streiner and Norman, 1995). Allen and Locker (1997), in their study using weighted and non-weighted versions of the Oral Health Impact Profile, found that weighted scores actually performed no better than simple summation of the scores.

Muldoon et al. (1998) proposed two operational definitions of quality of life: objective functioning and subjective well-being. All HRQL instruments purport to measure one or both of these but, to date, researchers have been reluctant to deal with the distinction between the two models. Most of the early measures were designed to assess function and this also applies to some of the contemporary measures. Others assess HRQL in a subjective manner; for example, how does the patient feel? Subjective appraisal is further complicated by the fact that it may also be influenced by psychological factors unrelated to health.

Theoretical models of HRQL have been the subject of much debate. Discussions have focused on definitions of HRQL, the dimensions which constitute HRQL and
appropriate HRQL scales. Debates have also involved the relative importance of multidimensional scales and single global concepts as well as the relative importance of generic versus condition-specific instruments (Cleary, 1996). HRQL is usually accepted as a multifactorial, lifelong continuous variable and it is, therefore, difficult to establish a baseline measurement. It is for this reason that longitudinal changes during illness or intervention are frequently the measurements of interest with individual patients acting as their own control (Schipper et al., 1996).

1.3 QUALITY OF LIFE AND HEALTH CARE OUTCOMES

Recent years have seen an increased interest in health related quality of life for three main reasons: firstly, to widen the range of outcome measures used in clinical trials and other evaluative studies; secondly, to allow auditing of the performance of individuals or groups; and thirdly, to monitor health levels in whole communities and provide a basis for developing policy (EuroQol Group, 1990; Jenkinson et al., 1993; Brown and Gordon, 1999).

When investigating acute medical conditions, measures such as mortality, symptom resolution and return to baseline health are accepted outcome measures. However, when considering chronic conditions, outcome measures become more difficult to establish and enhancement of HRQL is an important issue (Brown et al., 1999). Measuring medical outcomes has become even more important with current economic constraints, but how to do this remains a problem. The standardised self-report survey is emerging as the most appropriate method from the patient’s viewpoint (Ware, 1993).

Quality of life as an outcome indicator has been added to social and health service programme development in recent years. The Department of Health proposed that the following should be incorporated into outcome assessment: survival rates; symptoms and complications; health status and quality of life; the experiences of patients and
their carers; and the costs and use of resources (Bowling, 1995). The arguments in favour of measuring quality of life as an outcome of clinical interventions have been on the basis that the main goals of medical treatment are to increase survival and to add quality to that survival. Measurement of HRQL has been facilitated as evidence accrues to show that measures of HRQL are both valid and reliable (Wilson and Cleary, 1995).

1.4 METHODOLOGICAL ISSUES

There are many potential sources of variation in quality of life surveys and there is no general consensus at the present time on measurement. Allison et al. (1997) questioned whether quality of life is a dynamic construct. They proposed that the features of life that make the most important contributions to its quality do change. They stated the example of an individual who may describe family, work, finances and health as the most important aspects, but a year later following surgery for laryngeal cancer may describe family and being able to eat and speak as the most important aspects of life. Therefore, terms of reference may alter and when measuring changes in HRQL it is sometimes difficult to know whether these are absolute changes due to intervention, or whether the variation is due to altered reference standards (Allison et al., 1997). This problem area can be resolved to some extent by using QOL instruments with additional measures and also determining pre-treatment characteristics.

Hornberger and Lenert (1996) proposed that variations in quality of life measurements may arise due to one or more of the following:

- Variations in the theoretical framework for defining quality of life.

Three general theories have been used for characterising quality of life attributes: the Hedonic Theory in which the ultimate good life is certain conscious experiences such as pleasure, happiness or enjoyment; the Preference Satisfaction Theory which considers the importance one attaches to satisfaction of desires or preferences; and
the Ideal Theory which relates to the attainment of specific, explicitly normative ideals.

- Variations in methods for eliciting quality of life responses.
  This varies depending on the attributes included in the survey and the relative importance given to these attributes. Descriptions of health states may also have an effect (so-called framing effects), as will timing of the survey and the media used for conducting the survey.

- Respondent characteristics.
  Aspects such as sociodemographic factors and current health state may also influence quality of life.

Read (1993) discussed a number of other methodological issues involved in QOL research:

- How do traditional medical measures “fit in” with quality of life assessments?
  Experiments involving traditional medical variables remain important for understanding disease pathophysiology and mechanisms by which therapy acts. The newer HRQL measures assess the extent to which these therapies benefit the patient, in terms which are relevant to the individual. There is therefore a role for both approaches in modern medicine.

- Are quality of life measures more suited to use in patient care or should they be limited to use in clinical trials?
  It is generally considered that HRQL measures are more appropriate for evaluation studies than in individual patient care. Patient care often depends more on face-to-face judgements and interaction.

- How practical are these scales for use in clinical trials?
  Some of the scales which are available are more practical than others. Some of the very long questionnaires preclude them from use in the majority of studies. It remains a problem, however, that too few resources are expended in this area of research.
• Do quality of life instruments produce “hard” or “soft” data?
  It is often perceived that traditional medical research produces hard data and that
  studies such as quality of life research, which involve cognitive information,
  produce so-called “soft” data. However, many of the HRQL instruments perform
  better than traditional medical tests in terms of reliability, so this distinction
  between the two types of data does not really apply.

• Which are most useful: disease-specific measures or overall measures of health?
  This really depends on the purpose of the study and frequently the use of both
  types of instrument is recommended.

• Is the notion of quality of life too personal or individual to be reduced to
  standardised measures?
  People’s perceptions of quality of life differ. However, there is empirical data to
  support the argument that there remains a good deal of common ground among
  relative evaluation of health states.

1.5 QUALITY OF LIFE INSTRUMENTS

When measuring HRQL three questions must be asked (Dolan, 1999):

• What is to be valued? For example, the alternative types and levels of HRQL which
  the individual may experience during the period of interest.

• How is it to be valued? What measurement techniques will be used?

• Who is to value it? This is usually the individual who is in that health state.

The measurement of HRQL is far from easy and there are a number of issues which
continue to cause problems for researchers working in this field (Brown and Gordon,
1999). As mentioned previously, the first problem is that there is little agreement on
the definition of quality of life (Brown and Gordon, 1999; Dijkers, 1999). This is
compounded by the fact that many studies are carried out with little thought for the
most appropriate quality of life instrument and without establishing a research
question. Each instrument has a specific focus and level of sensitivity and the
appropriate choice needs careful thought. In addition, many instruments are cumbersome and more appropriate for research than in a clinical setting. It is also debatable as to what should be included and whether different sections should be weighted. The difficulties in developing new instruments are compounded by the fact that there is no gold standard with which to compare (EuroQol Group, 1990). Early instruments produced crude measures as they focused only on morbidity and mortality. Recent years have seen the development of many more instruments with some measuring specific aspects and others measuring “global” quality of life. HRQL instruments may be administered in a number of ways, all of which have varying strengths and weaknesses. Methods of administration include direct interview, telephone interview, self completion questionnaires and surrogate responders if the individual is unable to answer the questions themselves (Guyatt et al., 1993). The most popular method is, almost without exception, the patient-completed questionnaire.

<table>
<thead>
<tr>
<th>MODE OF ADMINISTRATION</th>
<th>STRENGTHS</th>
<th>WEAKNESSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer administered</td>
<td>• Maximise response rate&lt;br&gt;• Few, if any, missing items&lt;br&gt;• Minimise errors of misunderstanding</td>
<td>• Requires many resources&lt;br&gt;• May reduce willingness to acknowledge problems</td>
</tr>
<tr>
<td>Telephone administered</td>
<td>• Few, if any, missing items&lt;br&gt;• Minimise errors of misunderstanding&lt;br&gt;• Less resource intensive than interviewer administered</td>
<td>• Limits format of the instrument</td>
</tr>
<tr>
<td>Self- administered</td>
<td>• Minimal resources required</td>
<td>• Greater likelihood of missing items, non-response and misunderstanding</td>
</tr>
<tr>
<td>Surrogate responders</td>
<td>• Reduces stress for groups such as the very ill/old&lt;br&gt;• Can include patients unable to respond themselves</td>
<td>• Perceptions of surrogate may differ from the target group</td>
</tr>
</tbody>
</table>
Brown (1999) proposed a number of areas in which problems may occur and how these can be avoided:

- Choice of the correct instrument. Is it appropriate for the study group in question? Is a generic or condition-specific instrument required? Is the instrument sufficiently responsive?
- Timing of HRQL measurement. Ideally this should not be so close to the intervention that the patient confuses changes in their HRQL with effects of the intervention.
- Frequency of measurement. Is the measurement once only, cross-sectional or part of a longitudinal study?
- Ceiling effects (large numbers of respondents with the highest score) and floor effects (large numbers of respondents with the lowest score) can be a problem. These issues usually arise when using an instrument which was derived using a different group of individuals and is not really appropriate for the respondents in question.

HRQL instruments have been classified in many different ways:

- A HRQL instrument may be designed to be evaluative (measures how HRQL has changed) or discriminative (to differentiate between people with better or worse HRQL). The construction of instruments is different depending on which aspect the research is primarily concerned.
- The subscales of questionnaires may be of two types: *domain subscales* which are designed to be content valid and the categories are determined by observers or *construct subscales*, which are designed to be construct valid and are assumed to map constructs which exist in patients (Hyland *et al*., 1994).
- Instruments may be objective or subjective (Dijkers, 1999). The objective measures are based on the assumption that opinions and values as to what constitutes a good life are shared by most individuals, whereas subjective measures are based on the assumption that only the individual involved can judge his/her quality of life.
- HRQL measures may be divided into two further categories: those which use proxy indicators (for example, work-loss associated with dental problems) or those non-
proxy measures incorporated under the title of generic and condition-specific measures (Corson et al., 1999). Whilst the first method is relatively easy to use, the second gives a more accurate representation of the situation.

Table 41 Comparison of evaluative and discriminative instruments.

<table>
<thead>
<tr>
<th>EVALUATIVE</th>
<th>DISCRIMINATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Responsiveness is important</td>
<td>• Reliability is important</td>
</tr>
<tr>
<td>• Correlations of changes in measures during a period of time, consistent with theoretically derived predictions</td>
<td>• Correlations between measures at a point in time, consistent with theoretically derived predictions</td>
</tr>
<tr>
<td>• Differences within patients during a period of time can be interpreted as trivial, small, moderate or large</td>
<td>• Differences between patients at a point in time can be interpreted as trivial, small, moderate or large</td>
</tr>
</tbody>
</table>

(Guyatt et al., 1993)

There are two main groups of HRQL instruments. Both approaches have their strengths and weaknesses and there are advantages to using both instruments in a research study (Ware, 1993; Garratt et al., 1996a, 1996b):

- **Generic measures** provide a summary of HRQL and may generate a profile measure or a single index of health.
- **Condition-specific measures** focus on a particular condition, disease, population or problem and are devised to measure patients’ perceptions of the outcomes of health care interventions.
Table 42 Comparison of generic and condition-specific HRQL instruments.

<table>
<thead>
<tr>
<th></th>
<th>STRENGTHS</th>
<th>WEAKNESSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENERIC Health profile</td>
<td>• Single instrument</td>
<td>• May not focus adequately on area of interest</td>
</tr>
<tr>
<td></td>
<td>• Detects differential effects on different aspects of health status</td>
<td>• May not be responsive</td>
</tr>
<tr>
<td></td>
<td>• Comparisons across different interventions, conditions possible</td>
<td>• Difficulty determining utility values</td>
</tr>
<tr>
<td>Utility measurement</td>
<td>• Single number representing net impact on quantity and quality of life</td>
<td>• Does not allow examination of effect on different aspects of quality of life</td>
</tr>
<tr>
<td></td>
<td>• Cost-utility analysis possible</td>
<td>• May not be responsive</td>
</tr>
<tr>
<td>CONDITION-SPECIFIC</td>
<td>• Clinically sensible</td>
<td>• Does not allow cross-condition comparisons</td>
</tr>
<tr>
<td></td>
<td>• May be more responsive</td>
<td>• May be limited in terms of populations and interventions</td>
</tr>
</tbody>
</table>

1.5.1 Generic instruments
There are two main types of generic instruments (Camilleri-Brennan and Steele, 1999). The first is the health profile in which a separate score is given to each domain, for example, the Short Form-36 health survey (SF-36). The other type of generic measures are the health indices where scores generated from all answers are added up to give a single number or index, for example, the EuroQol (EuroQol Group, 1990). Generic measures have uses in comparisons across populations and have scope for use in economic evaluation but they show limited ability to capture the effects of certain interventions.
The MOS Short Form-36 health survey (SF-36)

The SF-36 was developed as part of the Medical Outcomes Study (MOS) which was carried out in Boston, Chicago and Los Angeles during 1986-87. The aim of the study was to enhance the methods used for the monitoring of patient outcomes in practice and in research settings (McHorney et al., 1993). It is a practical measure to use due to its short length and it may be used as a self-administration instrument, for telephone administration, or in an interview situation (Ware and Sherbourne, 1992). It has undergone extensive psychometric testing in many countries including the United States and the United Kingdom (Ware and Sherbourne, 1992; Brazier et al., 1992, 1993; Garratt et al., 1993, 1994).

The SF-36 measures eight health concepts, the number of items in each section are shown in parentheses: Physical Functioning (10); Role-Physical (4); Role-Emotional (3); Bodily Pain (2); Social Functioning (2); Mental Health (5); Vitality (4); General Health (5) and change in health (1). The SF-36 may also be divided into a Physical Health Component (physical functioning, role-physical, bodily pain and general health) and a Mental Health Component (vitality, social functioning, role-emotional and mental health) using a conversion equation (see Appendix 5).

A study by Brazier et al. (1992) was one of the first to assess the test-retest reliability of the scale, with good results. A later study, also by Brazier et al. (1993), compared the SF-36 and EuroQol questionnaires using 1,980 patients in the Sheffield area. There was evidence that the EuroQol was less sensitive both at the ceiling (low levels of perceived ill-health) and throughout the range of health states. However, the authors noted that there was reasonable agreement between the scores for the two scales and the benefits are that they are both self-administered and neither is a great burden to the respondent.

Garratt et al. (1993, 1994, 1996a) have studied the SF-36 in United Kingdom populations in some depth. Their first study tested the SF-36 on 900 members of the general public as well as on patients with one of four common conditions: low back...
pain; menorrhagia; suspected peptic ulcer and varicose veins (Garratt et al., 1993). The study confirmed the psychometric validity and reliability for a UK population and significant differences were found for all four conditions when compared with the general population group. Garratt et al. (1996a) administered the SF-36 and a condition-specific measure to patients suffering from varicose veins over a period of 12 months. They noted that because varicose veins are a relatively minor condition which does not cause health to deviate far from the normative data for the general population, a narrowly focused specific instrument in conjunction with the SF-36 was beneficial.

The EuroQol
The EuroQol was developed by a multidisciplinary team from five European centres. It was designed to be a simple device which could be used alongside other measures, to enable comparison of results obtained in different disease groups and different settings and countries (Kind, 1996). It embraces six dimensions or domains with two or three categories within each. In the past, one of the main problems has been that there is no gold standard with which to compare new instruments. This was one of the aims of the development of the EuroQol (EuroQol Group, 1990).

Other generic measures include the General Health Questionnaire (Goldberg and Williams, 1988), the Sickness Impact Profile (Bergner et al., 1981) and the Spitzer Quality of Life Index (Spitzer et al., 1981).
1.5.2 Specific instruments

Specific instruments may be divided into four groups (Camilleri-Brennan and Steele, 1999):

• Condition or disease-specific: see below;
• Domain specific: focus on one domain, for example, depression or anxiety;
• Population specific: focus on one population group;
• Symptom specific: focus on one type of symptom, for example, pain.

Condition-specific

These are the most commonly used of the specific measures. Condition-specific measures are used in clinical situations and their narrow focus means that they are potentially more responsive to small, but clinically important, changes in health. However, Bulpitt (1997) qualified this with the warning that it is difficult to differentiate between lack of sensitivity of an instrument and whether quality of life does not change much.

A number of condition-specific quality of life measures have been developed for conditions as diverse as inflammatory bowel disease (Guyatt et al., 1989), rhinoconjunctivitis (Juniper and Guyatt, 1991), epilepsy (Baker et al., 1993), varicose veins (Garratt et al., 1996a), endometriosis (Bodner et al., 1997), obesity (Mathias et al., 1997) stroke (Williams et al., 1999), gastroesophageal reflux (Colwell et al., 1999), chronic airflow limitation (Guyatt et al., 1999) and chronic liver disease (Younossi et al., 1999). A number of these conditions have certain similarities with orthognathic treatment in that the patients do not have a fixed physical deficit and traditional scales of functional disability are inappropriate although day to day living may still be restricted.
The methodology first proposed by Guyatt et al. (1989) has been used to develop a number of condition-specific instruments which have been shown to be practical, valid, reliable and responsive (Guyatt et al., 1989; Juniper and Guyatt, 1991; Garratt et al, 1996b). The methodology is based on:

- Item selection: the health-related items for inclusion in the questionnaire are selected from the literature and from unstructured interviews with clinicians and patients. These items are then entered into the item reduction questionnaire. There may be up to 100 items in the initial questionnaire.

- Item reduction: the items on the questionnaire which are chosen most frequently and rated as being most important are included in the definitive questionnaire.

Juniper et al. (1997) compared two different methods of selecting items for a disease specific questionnaire: the impact method (now the most commonly used method) selects items that are most frequently perceived as important by patients, and the psychometric method (factor analysis) selects items primarily according to their relationship with one another. The authors found that the impact method produced a 32 item instrument and factor analysis a 36 item instrument. Twenty items were common to both which led them to conclude that different approaches do produce different instruments.

Due to the nature of these condition-specific measures, many are developed on relatively small sample sizes and then tested on larger groups. For example, Veldhuyzen van Zanten et al. (1993) used 24 patients for the item reduction phase in their gastritis study and then evaluated the instrument on 55 patients before and one month after treatment. Hyland et al. (1994) developed the items for their breathing problems questionnaire on only 15 patients and undertook the Principal Component Analysis on 89 patients. Williams et al. (1999) used 32 patients to generate the items for their stroke-specific quality of life scale and then tested it on 72 patients in the main study. Younossi et al. (1999) used 75 patients for item reduction and 130 subjects for the main study but there were several different types of liver disease represented within this group. Brown et al. (1999) used only 38 patients in their study.
of disease-specific and generic outcomes in patients with long term back pain. Guyatt et al. (1999) had 89 patients in their study but these were further divided into two groups where 45 were undergoing treatment and 44 were not. Juniper et al. (1999) commented on the small sample size (40 subjects) in their asthma study but also drew attention to the fact that it produced valid and reliable results.

The majority of HRQL instruments use a Likert scale with between four and seven choices (Revicki and Leidy, 1998). The Likert scale is an example of an ordinal scale and has been shown to correlate well with other much more complicated scaling methods (Likert, 1932; Murphy and Likert, 1937; Edwards and Kenney, 1946; all cited in Revicki and Leidy, 1998). The number of items in instruments varies widely. As the number of items in an instrument increases, the contribution of each to the total variance decreases and this results in a reduction in bias associated with each item. However, as the number of items increases the demand on the respondent to complete the questionnaire is also increased.

Guyatt et al. (1989) and Berzon (1998) stated that certain criteria must be fulfilled by condition-specific instruments:

- Items must reflect areas of physical and emotional function that are important to patients with the condition.
- Summary scores should be amenable to statistical analysis.
- Repeated administration in stable patients must yield similar results.
- When even a small clinically important change in score has occurred, the questionnaire score should reflect that change.
- The questionnaire should be valid.
- The questionnaire should be relatively short and simple.

Guyatt (1997) also argued that investigators should tap into individual patient values otherwise they are measuring only health status not HRQL. The fact that many condition-specific measures use patient-derived questions makes them very useful from this viewpoint.
Missing data must be handled carefully in QOL studies. Missing data may result in bias and make results of a trial uninterpretable. For example, in a trial for a new drug, the missing data may be those individuals who have had adverse reactions. There are two types of missing data: item non-response where one question is left unanswered or unit non-response where a whole questionnaire is omitted (Curran et al., 1998; Camilleri-Brennan and Steele, 1999). On the whole, the former is more easily managed than the latter. The question which must be asked is "Are the characteristics of patients with missing data different from those for whom complete data are available?" Reasons behind missing data are varied. There may have been administrative or logistic problems, the patients may have been unable to understand or read the questionnaire or they may have failed to appreciate the aims of the study. There are a number of different methods available for dealing with missing data. These include: treating the whole scale as missing if one item is missing and accept that data are lost; simple mean imputation where the scale score is calculated from the mean of the items; and general imputation methods which have been developed to estimate the most likely value for the missing data (Curran et al., 1998).

1.6 ORAL HEALTH AND QUALITY OF LIFE

"Oral health is a standard of health of the oral and related tissues which enables an individual to eat, speak and socialise without active disease, discomfort or embarrassment and which contributes to general well-being."

Department of Health (1994)

The importance of quality of life has seen widespread acceptance in medicine but oral health status has only recently been seen in these terms. This is despite the fact that oral health was first considered in terms of quality of life during World War II when the presence of six opposing teeth was used as an indicator of oral functioning and well-being and was used to assess suitability for service (Hatch et al., 1998).
keynote quality of life textbook includes chapters relating to quality of life and oncology, mental health, paediatrics, coronary artery disease, geriatrics, epilepsy, gastrointestinal disease, renal disease, obstructive airway disease, asthma and allergies, diabetes, rheumatic disorders, orthopaedics, and eye disease but there is no mention of dentistry or oral health (Spilker, 1996). This is despite the fact that oral diseases and conditions are highly prevalent and the consequences are not only physical but also economic, social and psychological. Oral disease seriously impairs quality of life in a large number of individuals and may affect various aspects of life including function, appearance and interpersonal relationships (Gift and Redford, 1992).

Gift and Atchison (1995) presented one of the keynote papers in the field of oral health in which they aimed to improve the understanding of the interaction between oral health, oral health-related quality of life and systemic health. Gift and Atchison (1995) also stressed the need to conceptualise oral health as an integral part of general health and to consider the contribution of oral health to overall HRQL. Oral health-related quality of life encompasses multiple domains as for HRQL. These include survival of the individual and the dentition; absence of disease or symptoms; appropriate physical functioning as associated with chewing and swallowing; absence of discomfort or pain; emotional functioning associated with smiling; social functioning associated with normal roles; perceptions of excellent oral health; satisfaction with oral health and no social or cultural disadvantages due to oral status. These domains also show complex interrelationships.

Traditionally there has been a tendency to treat the oral cavity as an autonomous anatomical landmark which happens to be located within the body and, as such, the oral cavity has been seen as separate to the body and the individual. As mentioned in Chapter 2, this approach is now being questioned and this has given rise to new concepts in which oral conditions are linked to diseases in other body locations and to health outcomes of QOL (Locker, 1997).
With increasing numbers of older adults retaining their natural dentition, the impact of oral health on quality of life is likely to increase in significance (Gift and Redford, 1992). The need for a comprehensive approach to study the social and psychological impact of oral disease was first realised in the late 1980s. Reisine et al. (1989) used a battery of previously validated scales to determine the impact of several common but serious dental conditions on quality of life and their findings indicated that a number of patients were affected in their home, social and leisure activities as a result of their dental condition.

There is now a growing recognition of the importance of quality of life in the field of dentistry and this has led to the development of oral health-related quality of life instruments. However, there are still relatively few patient derived indices. Corson et al. (1999) summarised the most popular instruments:
The best known of these measures is the Oral Health Impact Profile or OHIP (Slade and Spencer, 1994; Slade, 1997, 1998). The original version of the scale comprised 49 items divided into seven domains but a recent study produced a short form OHIP containing only 14 items (Slade, 1997). The OHIP is designed for older patients, to

<table>
<thead>
<tr>
<th>INSTRUMENT</th>
<th>COUNTRY</th>
<th>DETAILS</th>
</tr>
</thead>
</table>
| Subjective Oral Health Status Indicators - SOHSI | Canada  | • Derived from conceptual model for measuring oral health  
|                                                 |         | • Battery of eight dimensions                                                                                                          |
| The Oral Health Impact Profile (OHIP)           | Australia | • Derived from a conceptual model for measuring oral health  
|                                                 |         | • Original format was a 49-item questionnaire  
|                                                 |         | • Short form with 14 items  
|                                                 |         | • Reference population used to weight items  
|                                                 |         | • Included in UK Adult Dental Health Survey  
|                                                 |         | • Variety of scoring methods                                                                                                           |
| Geriatric Oral Health Assessment Index (GOHAI)  | USA     | • Originally 36-item questionnaire  
|                                                 |         | • Short form with 12 items                                                                                                               |
| The Dental Impact of Daily Living (DIDL)        | UK      | • Five dimensions  
|                                                 |         | • 36-item questionnaire  
|                                                 |         | • Respondent assigns weights                                                                                                             |
| Dental Health Status Quality of Life Questionnaire (DS-QoL) | UK | • Derived from a generic measure  
|                                                 |         | • 5 dimensions  
|                                                 |         | • Capable of generating both profile and index                                                                                           |
| OIDP                                            | UK      | • Derived from conceptual model for measuring oral health  
|                                                 |         | • Assesses impact of oral disorders on 7 daily tasks  
|                                                 |         | • Frequency and severity of impact is calculated                                                                                         |
| Dental Impact Profile                           | USA     | • 25 items  
|                                                 |         | • Assesses positive and negative impacts  
|                                                 |         | • Aims to assess cultural influences on perceptions of oral health                                                                   |
determine perceptions of the social impact of oral disorders and it has been used widely, mainly on patients over 60 years of age. Other instruments include the Social Impacts of Dental Disease, which was one of the first socio-dental indicators (Cushing et al., 1986); the Geriatric Oral Health Assessment Index (Atchison, 1997) and the Dental Impact Profile (Strauss, 1997). All of which were initially developed for use with an older population. A comparison of the OHIP with the SF-36 in three groups of patients: edentulous patients seeking implants, edentulous patients seeking conventional dentures and dentate patients, found that the SF-36 failed to discriminate between the groups and all sub-scale scores were within the range of normative data for UK adults. The OHIP did, however, discriminate between the three groups and the authors postulated that it could be used as a means of identifying those patients who would benefit from implant treatment (Allen et al., 1999). Recent work by McGrath and Bedi (1999) also studied quality of life in older people. Their study of 454 adults over the age of 65 found that over 70% perceived oral health status as being important to their quality of life, with the greatest impact on function.

Hatch et al. (1998) compared the effects of rigid fixation and intermaxillary wire fixation on HRQL in patients undergoing orthognathic surgery. Surprisingly, there was no significant difference in HRQL between the two groups at any time period although both groups showed progressive and statistically significant improvements in HRQL following surgery.

A number of papers in the field of head and neck oncology have also been published (Rogers et al., 1999a, 1999b; Konstantinovic, 1999). Rogers et al. (1999a) studied papers looking at quality of life in patients with head and neck cancer and provided an overview of quality of life questionnaires used in oral cancer research. The paper stressed that although there is a great wealth of information on quality of life, there is very little information relating to head and neck oncology. This supports the findings that although quality of life in some fields of medicine is well developed, this is not so in the fields of dentistry and maxillofacial surgery.
There is little doubt that oral health is an integral part of general health and contributes to overall health-related quality of life. The oral cavity contributes to HRQL at both biological and social psychological levels. When oral health is compromised, overall health and HRQL may be adversely affected (Gift and Atchison, 1995).

1.7 QUALITY OF LIFE: THE FUTURE

There remain many areas in HRQL research which require more work. Cleary (1996) proposed four areas of immediate importance:

- Development and testing theoretical models of the associations among the types of clinical variables and measures of HRQL, as well as the various components of HRQL measures.
- Analysis and interpretation of multidimensional HRQL measures.
- Analytical models used to develop and evaluate HRQL.
- Better ways of calibrating changes in HRQL scales that are more meaningful to all concerned.
2.0 AIMS

Until recently, research into orthognathic treatment has tended to concentrate on subjective opinions and traditional measures of morbidity (for example, paraesthesia). The assessment of outcomes of orthognathic treatment is a difficult concept in that the patient is not ill prior to treatment and their life is not extended by treatment. For this reason, a number of the quality of life measures which are available are not appropriate for use with orthognathic patients.

The aims of this study were therefore to:

- develop a condition-specific quality of life measure for patients requesting orthognathic treatment, using standard techniques of item derivation and testing. The instrument will be known as the Orthognathic Quality of Life Questionnaire (OQLQ).
- assess the reliability, validity and responsiveness of the instrument.
- assess changes in the OQLQ and the SF-36 at three time periods during treatment.
3.0 MATERIALS AND METHODS

Ethical approval was obtained from the Joint Research and Ethics Committees of the units involved. All patients were provided with an information sheet and completed a consent form. Participants were assured that all responses were confidential and were advised that failure to participate in the study or withdrawal from the study during treatment would not affect their care in any way.

3.1 DEVELOPMENT OF THE INSTRUMENT

The development of the instrument was in three stages (Guyatt et al., 1989; Peto et al., 1995):

1. Item generation.
2. Item reduction.
3. Testing of the instrument.

3.1.1 Item generation

This stage involved producing a list of items, the content of which would reflect the impact of facial deformity on patients’ quality of life. Items were derived following a thorough literature review and interviews with health professionals and patients.

Unstructured interviews were conducted by the same interviewer (Miss S. Cunningham). A convenience sample of six maxillofacial surgeons (senior registrar or consultant level) and ten orthodontists (senior registrar or consultant level) from teaching hospitals and district general hospitals in the United Kingdom were interviewed either in person or by telephone. All clinicians interviewed were asked how they believed dentofacial deformities affected patients’ quality of life and their statements were listed.
Interviews were also conducted with ten patients who were intending to start orthognathic treatment and ten patients who were undergoing pre-surgical orthodontics. These individuals were selected, based on the next twenty consecutive patients seen in the Orthodontic department of the Eastman Dental Hospital who fulfilled the following selection criteria:

- over 16 years of age;
- spoke fluent English;
- suffered from no congenital deformity (for example, cleft lip/palate);
- intended to embark on (or were currently undergoing) a course of treatment involving both orthodontics and orthognathic surgery;
- were NOT involved in any other section of this PhD.

The interviews were undertaken away from the clinical area and lasted between 10 and 30 minutes depending on the amount of information interviewees were able to provide. All interviews were performed personally and patients were asked to "tell me about aspects of your everyday life which are affected in any way by concerns regarding your dental appearance and/or facial appearance and/or bite". All statements were listed as verbatim comments along with those from the health professionals. This produced a list of 56 items which were then scrutinised for ambiguity and repetition, leaving 42 distinct items.

### 3.1.2 Item reduction

The second stage, item reduction, assessed the response frequency by respondents to each item. The list of items was administered to a group of 46 patients who were not involved in the item generation stage or in any other section of this PhD, 44 were returned (96% response rate). Of these respondents, 30 completed the questionnaire prior to any treatment and the remaining 16 were undergoing pre-surgical orthodontics. Respondents were asked to mark every item which applied to them and which was relevant in their decision to proceed with orthognathic treatment.
Items which were selected by 20% or less of the respondents were excluded from the final measure, based on the work undertaken by Streiner and Norman (1995), which proposed that items not frequently chosen are irrelevant to the majority of patients. This method of item reduction produced a list of 22 statements which were used in the final instrument. The results of item reduction are shown in Table 44.

3.1.3 Testing the instrument

The final instrument was pre-tested on five patients. These individuals were asked to complete the questionnaire in their own time and were then asked to explain what they understood by each item. They were also asked to comment on any problems with wording or understanding and a number of minor changes in wording were made following this.

Evaluation of the measure took place within an existing longitudinal study of orthognathic patients in three orthodontic/maxillofacial units in the South East of England (see Chapter 4). Eighty-eight patients were recruited for the study and a total of 85 replies were received (response rate 97%).

The 22 items were presented in the order they had been recorded during interviews and were not in any definite arrangement at this stage (Appendix 4). Subjects were asked to grade the 22 statements based on a four point scale where 1 meant the issue covered in the statement bothered them a little, 4 meant it bothered them a lot and 2/3 was between. An option of N/A (not applicable) existed for those not affected by the issue or if the issue in the statement did not bother them. The number of response options was based on previous research which showed the optimum number of responses is between 3 and 7 (Hyland et al., 1991). The total score ranged from 0 to 88, with lower scores indicating better quality of life and higher scores indicating poorer quality of life. A principal component analysis (PCA) utilised this data to establish the main domains (see section 3.5 Statistical analysis).
3.1.4 Timing of questionnaires

A longitudinal study was also required to assess validity and responsiveness. It was estimated that the first 65 respondents to be recruited should complete treatment within the time constraints of the research. Therefore, these subjects were asked to complete questionnaires on three occasions:

T1 Before any active treatment started;
T2 Following pre-surgical orthodontics/prior to surgery (at least 6 weeks prior to surgery);
T3 Four to six weeks following debond.

A 100mm Visual Analogue Scale (VAS), labelled from 0-10, was also included at each time period. This asked respondents to rate how they felt about their dental and facial appearance and oral function, with 0 representing “no problem at all” and 10 being “the worst problem imaginable”. The Short Form-36 outcome measure was also included at T2 and T3 (see Appendix 5). The SF-36 was chosen because it is the most widely used patient assessed measure of health outcome and there is good evidence of validity, reliability and responsiveness in a wide range of different populations (Ware and Sherbourne, 1992; Brazier et al., 1993; Ware, 1993; Garratt et al., 1994, 1996a; Kaplan, 1998). In addition, the SF-36 has been used in the past for purposes of assessing validity and responsiveness (Garratt et al., 1996a).

3.1.5 Response rates

Eighty-eight questionnaires were distributed at T1 and 85 returned (97% response rate). The responses at T1 were used for the PCA. In the time frame of the study, 65 patients reached the T2 stage and 62 questionnaires were returned (95% response rate). The T2 data was used for validity testing. The first 30 cases with complete data at T1, T2 and T3 were used for responsiveness testing.


3.2 RELIABILITY TESTING

Reliability was assessed using two methods:

- Test-retest: see below.

- Assessment of internal consistency of the separate domains: this involved testing for homogeneity and measured correlations between items in each domain.

3.2.1 Test-retest

Of the 85 respondents at T1, 25 were asked to repeat the same questionnaire four to six weeks after the first in order to assess test-retest reliability. This sample size was selected in accordance with work by Guyatt et al. (1987). The time interval of 4-6 weeks was selected because the respondents had not started treatment and therefore the assessment should have remained stable. In addition, they were then unlikely to remember how they answered the first time. The VAS was also included in the retest although the SF-36 was not, as it was felt that repeatability for this questionnaire had been tested sufficiently in the past. Twenty-four respondents returned both questionnaires (response rate 96%).

3.3 VALIDITY TESTING

Validity of the four separate domains was undertaken using the data obtained at T2 (following pre-surgical orthodontics/prior to surgery). T2 was selected as all instruments were administered at that time point.

In the absence of a gold standard of the impact of dentofacial deformity on patients’ quality of life, validity was assessed through comparisons with measures that were expected to be related to OQLQ scores. These were the 100mm VAS and the SF-36. Correlations were derived based on a thorough knowledge of the literature and on clinical judgement.
The following associations were hypothesised:

- A moderate positive correlation would exist between all four domains of the OQLQ and the 100mm visual analogue scale.
- A moderate negative correlation would exist between the OQLQ domains and the SF-36 Mental Health Component Summary (MCS) Score and poor correlations with the SF-36 Physical Health Component Summary (PCS) Score.
- Moderate negative correlations would exist between the social aspects domain/dentofacial aesthetics domain and the SF-36 scales of Social Functioning, Mental Health, Role-Emotional, General Health and Vitality.
- A moderate negative correlation would exist between the oral function domain and the SF-36 scale of Social Functioning.
- A moderate negative correlation would exist between the awareness of dentofacial aesthetics domain and the SF-36 Mental Health scale.
- Poor levels of correlation would exist between all domains of the OQLQ and the SF-36 scales of Physical Functioning, Role-Physical and Bodily Pain.

3.4 RESPONSIVENESS TESTING

If the OQLQ is to be used in an evaluative context then it must be responsive to small but important changes in health. The responsiveness of the OQLQ was assessed by comparing scores before and after treatment. Responsiveness was assessed using the first 30 completed cases. Sample size was in accordance with a number of previous studies including work by Guyatt et al. (1987) who used 13 patients in one study and 28 in another.

A health transition question was also included in the T3 questionnaire to aid responsiveness testing. This question asked respondents to indicate whether their dental and facial problems were: much better; better; the same as; worse or much worse than one year previously.
3.5 STATISTICAL ANALYSIS

Statistical analysis was undertaken using SPSS for Windows\textsuperscript{1}.

3.5.1 Instrument testing (T1 data)

**T1 Data: Assessment of response frequencies**

Response frequencies were initially assessed to ensure that a substantial number of respondents were not answering the same way for any one item as such questions then lack discriminatory power. No single item had a response frequency over 80%, therefore, it was not necessary to exclude any items at this stage.

**Principal component analysis (PCA)**

In order to establish that a satisfactory factor analysis could be undertaken, it was first established that the Kaiser-Meyer-Olkin measure of sampling adequacy was greater than 0.5 and that the Bartlett test of sphericity was significant. The eigenvalue was set at 1.0. The relationship between the variables and some of the factors is maximised by rotation which makes interpretation of the factor analysis easier and the method chosen was varimax rotation. Loadings with values less than 0.4 were suppressed. Factor analysis then groups items with high correlations into factors (or domains as they are frequently known). The first few principal components account for most of the variability of the original data, so that the remaining principal components can be disregarded and therefore reduce the dimensionality of the data (Petrie and Watson, 1999).

\textsuperscript{1} SPSS for Windows. SPSS UK Ltd. 1\textsuperscript{st} Floor St. Andrew's House, West Street, Woking, Surrey, GU21 1EP
3.5.2 Reliability testing

Assessment of internal consistency of the separate domains
All 85 responses from T1 were utilised in the internal consistency analysis. This involved testing for homogeneity and measured alpha coefficients and item-total correlations for each domain.

Test-Retest Reliability
Reliability (test-retest) was analysed using the methodology previously described in Chapter 1.

3.5.3 Validity testing
Spearman correlations were calculated for the relationships between the OQLQ and VAS values at T2 and between the OQLQ and SF-36 components at T2. In previous work, Juniper et al. (1996) suggested the following levels of correlation should be used: 0.2-0.35; 0.35-0.5 and greater than 0.5 representing poor, moderate and strong correlations, respectively. These values were adhered to in the present study and were felt to be appropriate, particularly for the SF-36 which is a general health measure and has little or no application in dentistry. Hence it is difficult to specify, a priori, expected levels of correlation.

3.5.4 Responsiveness testing
The change in scores for the OQLQ and the VAS were calculated between T2/T1 and T3/T2. The change in SF-36 component scores was calculated between T3/T2 only. Standardised Response Means were then calculated and used to assess the responsiveness of the OQLQ, VAS and SF-36. The SRM is equal to the mean change
in score divided by the standard deviation of the change in score and allows meaningful comparisons between different instruments. SRMs of 0.2, 0.5 and 0.8 represent small, moderate and large changes respectively (Liang et al., 1990; Garratt et al., 1994).
4.0 RESULTS

4.1 SUMMARY DATA

At T1, 85 questionnaires were returned. Of these respondents, 53 were female and 32 male. The mean age was 22.4 years (95% CI 21.0 to 23.8 years). At T2, 39 respondents were female, 23 male and the mean age was 21.9 years (95% CI 20.3 to 23.5 years). Of the 30 respondents used for the responsiveness testing, 17 were female, 13 male and there was a mean age of 23.2 years (95% CI 20.5 to 25.8 years).

4.2 ITEM REDUCTION AND PCA

Table 44 shows all 44 items which were initially included in the instrument. The right hand column shows the number (and percentage) of respondents who stated that the item bothered them. The resultant OQLQ was produced from the 22 items which were selected most frequently and these are highlighted in the table.

The results of the PCA from these 22 items are shown in Table 45 and 46. A four component solution provided the most interpretable results:

1\textsuperscript{st} component or domain: Social aspects.

2\textsuperscript{nd} component or domain: Dentofacial aesthetics.

3\textsuperscript{rd} component or domain: Oral function.

4\textsuperscript{th} component or domain: Awareness of dentofacial aesthetics.
Table 44  
Item reduction: Items which were retained in the instrument at the item reduction stage are shown in bold.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Response frequency (n=44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I am self-conscious about the appearance of my teeth</td>
<td>38 (86%)</td>
</tr>
<tr>
<td>2. I have problems biting</td>
<td>30 (68%)</td>
</tr>
<tr>
<td>3. I have problems chewing</td>
<td>14 (32%)</td>
</tr>
<tr>
<td>4. There are some foods I avoid eating because the way my teeth meet makes it difficult</td>
<td>16 (36%)</td>
</tr>
<tr>
<td>5. I don't like eating in public places</td>
<td>16 (36%)</td>
</tr>
<tr>
<td>6. I get pains in my face or jaw</td>
<td>14 (32%)</td>
</tr>
<tr>
<td>7. I don't like seeing a side view of my face (profile)</td>
<td>30 (68%)</td>
</tr>
<tr>
<td>8. I avoid looking in mirrors</td>
<td>8 (18%)</td>
</tr>
<tr>
<td>9. I spend a lot of time studying my face in the mirror</td>
<td>14 (32%)</td>
</tr>
<tr>
<td>10. I spend a lot of time studying my teeth in the mirror</td>
<td>26 (59%)</td>
</tr>
<tr>
<td>11. I dislike having my photograph taken</td>
<td>38 (86%)</td>
</tr>
<tr>
<td>12. I dislike being seen on video</td>
<td>32 (73%)</td>
</tr>
<tr>
<td>13. I often stare at other people's teeth</td>
<td>20 (45%)</td>
</tr>
<tr>
<td>14. I often stare at other people's faces</td>
<td>14 (32%)</td>
</tr>
<tr>
<td>15. I am self-conscious about my facial appearance</td>
<td>36 (82%)</td>
</tr>
<tr>
<td>16. I try to cover my mouth when I meet people for the first time</td>
<td>16 (36%)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>17. I try to cover my face when I meet people for the first time</td>
<td></td>
</tr>
<tr>
<td>18. I worry about meeting people for the first time</td>
<td></td>
</tr>
<tr>
<td>19. I worry that people will make hurtful comments about my appearance</td>
<td></td>
</tr>
<tr>
<td>20. I am afraid that people will stare at me when I meet them for the first time</td>
<td></td>
</tr>
<tr>
<td>21. I think people underestimate my abilities because of the way I look</td>
<td></td>
</tr>
<tr>
<td>22. I lack confidence in most situations</td>
<td></td>
</tr>
<tr>
<td>23. I don’t like smiling when I meet people</td>
<td></td>
</tr>
<tr>
<td>24. I sometimes get depressed about my appearance</td>
<td></td>
</tr>
<tr>
<td>25. I worry about making friends</td>
<td></td>
</tr>
<tr>
<td>26. I worry that I will not get boyfriends/girlfriends</td>
<td></td>
</tr>
<tr>
<td>27. I worry that people will tease me or call me names</td>
<td></td>
</tr>
<tr>
<td>28. I feel that I am not as good as other people because I look a bit different</td>
<td></td>
</tr>
<tr>
<td>29. I don’t make friends very easily</td>
<td></td>
</tr>
<tr>
<td>30. I sometimes think that people are staring at me</td>
<td></td>
</tr>
<tr>
<td>31. I don’t like being in large groups because I feel self-conscious</td>
<td></td>
</tr>
<tr>
<td>32. I’m frightened that I will be rejected by boyfriends/girlfriends</td>
<td></td>
</tr>
<tr>
<td>33. I’m frightened that I will be rejected by friends</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>34. I think my facial appearance makes me look masculine (for females)</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>I think my facial appearance makes me look feminine (for males)</td>
<td></td>
</tr>
<tr>
<td>35. I get offended by things people say about my teeth/face</td>
<td>6 (14%)</td>
</tr>
<tr>
<td>36. I lack confidence at work</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>37. I think my appearance affects my chances of getting a good job</td>
<td>8 (18%)</td>
</tr>
<tr>
<td>38. I avoid going out to public places if possible</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>39. I’m afraid that no-one will want to marry me as I look at the moment</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>40. I’m afraid I will not get the job I want because of my facial appearance</td>
<td>2 (5%)</td>
</tr>
<tr>
<td>41. I’m afraid I will not get the job I want because of my lack of self-confidence</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>42. People make jokes about my appearance</td>
<td>7 (16%)</td>
</tr>
<tr>
<td>43. People make nasty comments about my appearance</td>
<td>8 (18%)</td>
</tr>
<tr>
<td>44. Comments about my appearance really upset me, even when I know people are only joking</td>
<td>18 (41%)</td>
</tr>
<tr>
<td>Table 45</td>
<td>Principal component analysis (PCA) data.</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>n=85</td>
<td>Initial eigenvalues/extraction sums of squared loadings</td>
</tr>
<tr>
<td>Component/domain</td>
<td>Total</td>
</tr>
<tr>
<td><strong>DOMAIN 1: Social aspects</strong></td>
<td></td>
</tr>
<tr>
<td>15. cover mouth when meeting people</td>
<td></td>
</tr>
<tr>
<td>16. worry about meeting people</td>
<td></td>
</tr>
<tr>
<td>17. worry people will make hurtful comments</td>
<td></td>
</tr>
<tr>
<td>18. lack confidence socially</td>
<td></td>
</tr>
<tr>
<td>19. do not like smiling</td>
<td></td>
</tr>
<tr>
<td>20. get depressed about appearance</td>
<td></td>
</tr>
<tr>
<td>21. sometimes think people are staring</td>
<td></td>
</tr>
<tr>
<td>22. comments about appearance upset me</td>
<td></td>
</tr>
<tr>
<td><strong>DOMAIN 2: Dentofacial aesthetics</strong></td>
<td></td>
</tr>
<tr>
<td>1. self-conscious about appearance of my teeth</td>
<td>3.13</td>
</tr>
<tr>
<td>7. don’t like seeing side view of face (profile)</td>
<td></td>
</tr>
<tr>
<td>10. dislike having photograph taken</td>
<td></td>
</tr>
<tr>
<td>11. dislike being seen on video</td>
<td></td>
</tr>
<tr>
<td>14. self-conscious about appearance</td>
<td></td>
</tr>
<tr>
<td><strong>DOMAIN 3: Oral function</strong></td>
<td></td>
</tr>
<tr>
<td>2. problems biting</td>
<td>1.33</td>
</tr>
<tr>
<td>3. problems chewing</td>
<td></td>
</tr>
<tr>
<td>4. avoid eating some foods</td>
<td></td>
</tr>
<tr>
<td>5. don’t like eating in public</td>
<td></td>
</tr>
<tr>
<td>6. pains in face/ jaw</td>
<td></td>
</tr>
<tr>
<td><strong>DOMAIN 4: Awareness of dentofacial aesthetics</strong></td>
<td></td>
</tr>
<tr>
<td>8. spend time studying face</td>
<td>1.13</td>
</tr>
<tr>
<td>9. spend time studying teeth</td>
<td></td>
</tr>
<tr>
<td>12. stare at people’s teeth</td>
<td></td>
</tr>
<tr>
<td>13. stare at people’s faces</td>
<td></td>
</tr>
</tbody>
</table>
Table 46  The rotated factor matrix.

<table>
<thead>
<tr>
<th>Component/Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td></td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td></td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td></td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td></td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td></td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td></td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8</td>
<td></td>
<td></td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td></td>
<td></td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Q10</td>
<td></td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11</td>
<td></td>
<td>0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12</td>
<td></td>
<td></td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>Q13</td>
<td></td>
<td></td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Q14</td>
<td></td>
<td></td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Q15</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q16</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q17</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q18</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q19</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q20</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q21</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q22</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NB: Rotation converged in 7 iterations.
4.3 INITIAL (T1) DATA

Tables 47 and 48 show the T1 data provided by all 85 respondents. Table 47 includes mean scores for each domain and also dimension scores (mean score divided by number of items) which allows comparison between domains with different numbers of items. Of note is the dentofacial aesthetic domain which showed the highest dimension score.

Table 48 presents the findings with respect to gender, females showed significantly higher scores than males for the social aspects domain and the dentofacial aspects domain, but no differences for oral function or awareness of dentofacial aesthetics.

Table 47  T1 data.

<table>
<thead>
<tr>
<th>Domain 1: Social aspects</th>
<th>Mean (95% CI)</th>
<th>Dimension score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 2: Dentofacial aesthetics</td>
<td>13.9 (12.7 to 15.1)</td>
<td>2.78</td>
</tr>
<tr>
<td>Domain 3: Oral function</td>
<td>8.4 (7.2 to 9.6)</td>
<td>1.68</td>
</tr>
<tr>
<td>Domain 4: Awareness of dentofacial aesthetics</td>
<td>7.6 (6.5 to 8.7)</td>
<td>1.90</td>
</tr>
</tbody>
</table>

Dimension score = total score divided by number of items
Table 48  Gender differences for domain scores.

<table>
<thead>
<tr>
<th>Domain</th>
<th>U value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 1: Social aspects</td>
<td>547.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Domain 2: Dentofacial aesthetics</td>
<td>565.50</td>
<td>0.01</td>
</tr>
<tr>
<td>Domain 3: Oral function</td>
<td>821.50</td>
<td>0.92</td>
</tr>
<tr>
<td>Domain 4: Awareness of dentofacial aesthetics</td>
<td>642.50</td>
<td>0.08</td>
</tr>
</tbody>
</table>

U=Mann-Whitney U value

4.4 RELIABILITY TESTING

Table 49 shows the data for internal consistency (item-total correlations and alpha coefficients). Table 50 shows the test-retest data. Both internal reliability and test-retest reliability provided results within the accepted range.
Table 49  Reliability testing: internal consistency.

<table>
<thead>
<tr>
<th>Component / domain</th>
<th>Item-total correl*</th>
<th>Alpha coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DOMAIN 1: Social aspects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. cover mouth when meeting people</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>16. worry about meeting people</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>17. worry people will make hurtful comments</td>
<td>0.88</td>
<td>0.93</td>
</tr>
<tr>
<td>18. lack confidence socially</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>19. do not like smiling</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>20. get depressed about appearance</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>21. sometimes think people are staring</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>22. comments about appearance upset me</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td><strong>DOMAIN 2: Dentofacial aesthetics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. self-conscious about appearance of my teeth</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>7. don’t like seeing side view of face (profile)</td>
<td>0.82</td>
<td>0.86</td>
</tr>
<tr>
<td>10. dislike having photograph taken</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>11. dislike being seen on video</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>14. self-conscious about appearance</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td><strong>DOMAIN 3: Oral function</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. problems biting</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>3. problems chewing</td>
<td>0.79</td>
<td>0.83</td>
</tr>
<tr>
<td>4. avoid eating some foods</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>5. don’t like eating in public</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>6. pains in face/jaw</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td><strong>DOMAIN 4: Awareness of dentofacial aesthetics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. spend time studying face</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>9. spend time studying teeth</td>
<td>0.67</td>
<td>0.87</td>
</tr>
<tr>
<td>12. stare at people’s teeth</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>13. stare at people’s faces</td>
<td>0.81</td>
<td></td>
</tr>
</tbody>
</table>
Table 50  Reliability testing: test-retest data.

<table>
<thead>
<tr>
<th></th>
<th>n=22</th>
<th>Pearson correlation</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual analogue scale</td>
<td></td>
<td>0.77**</td>
<td>0.75</td>
</tr>
<tr>
<td>Domain 1: Social aspects</td>
<td></td>
<td>0.88**</td>
<td>0.88</td>
</tr>
<tr>
<td>Domain 2: Dentofacial aesthetics</td>
<td></td>
<td>0.92**</td>
<td>0.87</td>
</tr>
<tr>
<td>Domain 3: Oral function</td>
<td></td>
<td>0.76**</td>
<td>0.74</td>
</tr>
<tr>
<td>Domain 4: Awareness of dentofacial aesthetics</td>
<td></td>
<td>0.89**</td>
<td>0.87</td>
</tr>
</tbody>
</table>

** p<0.01
4.5   VALIDITY AND RESPONSIVENESS TESTING

Spearman correlations between the OQLQ and the VAS and SF-36 are shown in Table 51. Comparison of these findings with the hypothesised correlations are shown in Tables 52 (hypothesised moderate correlations) and 53 (hypothesised poor correlations). Whether or not the hypothesis was confirmed is noted in the far right hand column. A high proportion of hypotheses were confirmed.

Table 54 reports the raw data for those individuals in the responsiveness testing group. This data is also graphically represented in Figure 35, where it can be noted that there were significant changes between T3/T2 for the social aspects domain (p<0.01), the dentofacial aesthetics domain (p<0.001) and the oral function domain (p<0.001). There were no significant findings between T2/T1. Table 55 shows the mean changes and SRMs for the VAS and OQLQ and Table 56 reports those for the SF-36. Both the VAS and OQLQ showed much larger SRMs at T3/T2 than T2/T1 and, in addition, the VAS and OQLQ showed much greater SRMs than the SF-36.

Data for the SF-36 at T2 and T3 are shown in Table 57 and Figure 36. These data are compared with those from the Oxford study (Jenkinson et al., 1993) in Table 58 and Figure 37.
Table 51  Validity testing: Spearman correlations between the condition-specific measure and the VAS and SF-36 scales at T2.

<table>
<thead>
<tr>
<th>n=62</th>
<th>OQLQ DOMAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Domain 1: Social aspects</td>
</tr>
<tr>
<td>VAS</td>
<td>0.45**</td>
</tr>
<tr>
<td>SF-36 COMPONENTS</td>
<td></td>
</tr>
<tr>
<td>Physical Functioning (PF)</td>
<td>-0.10</td>
</tr>
<tr>
<td>Role-Physical (RP)</td>
<td>-0.08</td>
</tr>
<tr>
<td>Bodily Pain (BP)</td>
<td>-0.07</td>
</tr>
<tr>
<td>General Health (GH)</td>
<td>-0.26*</td>
</tr>
<tr>
<td>Vitality (VT)</td>
<td>-0.40**</td>
</tr>
<tr>
<td>Social Functioning (SF)</td>
<td>-0.38**</td>
</tr>
<tr>
<td>Role-Emotional (RE)</td>
<td>-0.22</td>
</tr>
<tr>
<td>Mental Health (MH)</td>
<td>-0.47**</td>
</tr>
<tr>
<td>SF-36 Physical Component (PCS)</td>
<td>0.16</td>
</tr>
<tr>
<td>SF-36 Mental Component (MCS)</td>
<td>-0.50**</td>
</tr>
</tbody>
</table>

*p<0.05  **p<0.01
Table 52 Comparison of hypothesised and actual findings for hypothesised moderate correlations (0.35-0.5).

<table>
<thead>
<tr>
<th>OQLQ domain</th>
<th>Scale for comparison</th>
<th>Correlation</th>
<th>Hypothesis proved?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 1: Social aspects</td>
<td>Visual analogue scale</td>
<td>+0.45</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Mental health component score (MCS)</td>
<td>-0.50</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Social Functioning</td>
<td>-0.38</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Mental Health</td>
<td>-0.47</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Role-Emotional</td>
<td>-0.22</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>SF-36 General Health</td>
<td>-0.26</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>SF-36 Vitality</td>
<td>-0.40</td>
<td>Yes</td>
</tr>
<tr>
<td>Domain 2: Dentofacial aesthetics</td>
<td>Visual analogue scale</td>
<td>+0.44</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Mental health component score (MCS)</td>
<td>-0.43</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Social Functioning</td>
<td>-0.48</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Mental Health</td>
<td>-0.46</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Role-Emotional</td>
<td>-0.22</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>SF-36 General Health</td>
<td>-0.40</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Vitality</td>
<td>-0.34</td>
<td>Borderline</td>
</tr>
<tr>
<td>Domain 3: Oral function</td>
<td>Visual analogue scale</td>
<td>+0.34</td>
<td>Borderline</td>
</tr>
<tr>
<td></td>
<td>SF-36 Mental health component score (MCS)</td>
<td>-0.06</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>SF-36 Social Functioning</td>
<td>-0.11</td>
<td>No</td>
</tr>
<tr>
<td>Domain 4: Awareness of dentofacial aesthetics</td>
<td>Visual analogue scale</td>
<td>+0.27</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>SF-36 Mental health component score (MCS)</td>
<td>-0.35</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Mental Health</td>
<td>-0.29</td>
<td>No</td>
</tr>
</tbody>
</table>
Table 53  Comparison of hypothesised and actual findings for hypothesised poor correlations (below 0.35).

<table>
<thead>
<tr>
<th>OQLQ domain</th>
<th>Scale for comparison</th>
<th>Correlation</th>
<th>Hypothesis proved?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1:</strong> Social aspects</td>
<td>SF-36 Physical health component score (PCS)</td>
<td>+0.16</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Physical Functioning</td>
<td>-0.10</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Role-Physical</td>
<td>-0.08</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Bodily Pain</td>
<td>-0.07</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Domain 2:</strong> Dentofacial aesthetics</td>
<td>SF-36 Physical health component score (PCS)</td>
<td>-0.13</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Physical Functioning</td>
<td>-0.09</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Role-Physical</td>
<td>-0.14</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Bodily Pain</td>
<td>-0.18</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Domain 3:</strong> Oral function</td>
<td>SF-36 Physical health component score (PCS)</td>
<td>-0.23</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Physical Functioning</td>
<td>-0.23</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Role-Physical</td>
<td>+0.07</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Bodily Pain</td>
<td>-0.17</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Domain 4:</strong> Awareness of dentofacial aesthetics</td>
<td>SF-36 Physical health component score (PCS)</td>
<td>+0.32</td>
<td>Borderline</td>
</tr>
<tr>
<td></td>
<td>SF-36 Physical Functioning</td>
<td>+0.13</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Role-Physical</td>
<td>-0.13</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>SF-36 Bodily Pain</td>
<td>+0.19</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 54  Data for the responsiveness testing group.

<table>
<thead>
<tr>
<th>n=30</th>
<th>Start of treatment (T1)</th>
<th>Pre-surgery (T2)</th>
<th>End of treatment (T3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD) 95% CI</td>
<td>Mean (SD) 95% CI</td>
<td>Mean (SD) 95% CI</td>
</tr>
<tr>
<td>Domain 1: Social aspects</td>
<td>15.07 (10.39) 11.19 to 18.95</td>
<td>15.30 (9.99) 11.57 to 19.03</td>
<td>9.30 (8.14) 6.26 to 12.34</td>
</tr>
<tr>
<td>Domain 2: Dentofacial aesthetics</td>
<td>13.27 (5.92) 11.06 to 15.48</td>
<td>14.40 (4.65) 12.66 to 16.14</td>
<td>8.30 (4.68) 6.55 to 10.05</td>
</tr>
<tr>
<td>Domain 3: Oral function</td>
<td>8.23 (5.51) 6.17 to 10.29</td>
<td>10.03 (5.13) 8.11 to 11.95</td>
<td>6.50 (4.03) 5.00 to 8.01</td>
</tr>
<tr>
<td>Domain 4: Awareness of dentofacial aesthetics</td>
<td>7.20 (5.40) 5.18 to 9.22</td>
<td>7.80 (4.22) 6.22 to 9.38</td>
<td>7.03 (4.16) 5.48 to 8.58</td>
</tr>
</tbody>
</table>
Table 55  Responsiveness testing: Data for changes in scores between T2/T1 and T3/T2 for the OQLQ and VAS.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Mean score (SD) at baseline</th>
<th>Mean score (SD) at follow-up</th>
<th>Mean change (SD)</th>
<th>SRM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VAS (T2/T1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.15 (2.03)</td>
<td>5.86 (2.02)</td>
<td>-0.29 (2.29)</td>
<td>-0.13</td>
</tr>
<tr>
<td><strong>VAS (T3/T2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.86 (2.02)</td>
<td>1.47 (1.13)</td>
<td>-4.39 (2.01)***</td>
<td>-2.18</td>
</tr>
<tr>
<td><strong>OQLQ (T2/T1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social aspects</td>
<td>15.07 (10.39)</td>
<td>15.30 (9.99)</td>
<td>0.23 (6.89)</td>
<td>0.03</td>
</tr>
<tr>
<td>Dentofacial aesthetics</td>
<td>13.27 (5.92)</td>
<td>14.40 (4.65)</td>
<td>1.13 (3.72)</td>
<td>0.30</td>
</tr>
<tr>
<td>Oral function</td>
<td>8.23 (5.51)</td>
<td>10.03 (5.13)</td>
<td>1.80 (5.75)</td>
<td>0.31</td>
</tr>
<tr>
<td>Awareness of dentofacial aesthetics</td>
<td>7.20 (5.40)</td>
<td>7.80 (4.22)</td>
<td>0.60 (3.01)</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>OQLQ (T3/T2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social aspects</td>
<td>15.30 (9.99)</td>
<td>9.30 (8.14)</td>
<td>-6.00 (8.41)**</td>
<td>-0.71</td>
</tr>
<tr>
<td>Dentofacial aesthetics</td>
<td>14.40 (4.65)</td>
<td>8.30 (4.68)</td>
<td>-6.10 (5.41)***</td>
<td>-1.13</td>
</tr>
<tr>
<td>Oral function</td>
<td>10.03 (5.13)</td>
<td>6.50 (4.03)</td>
<td>-3.53 (4.31)***</td>
<td>-0.82</td>
</tr>
<tr>
<td>Awareness of dentofacial aesthetics</td>
<td>7.80 (4.22)</td>
<td>7.03 (4.16)</td>
<td>-0.77 (4.61)</td>
<td>-0.17</td>
</tr>
</tbody>
</table>

**p<0.01; ***p<0.001
Table 56  Responsiveness testing: Data for changes in scores between T3/T2 for the SF-36.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Mean score (SD) at baseline</th>
<th>Mean score (SD) at follow-up</th>
<th>Mean change T3/T2 (SD)</th>
<th>SRM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Functioning</td>
<td>95.00% (12.93%)</td>
<td>96.00% (6.21%)</td>
<td>1.00% (11.92%)</td>
<td>0.08</td>
</tr>
<tr>
<td>Role-Physical</td>
<td>92.50% (20.92%)</td>
<td>90.83% (26.65%)</td>
<td>-1.67% (30.75%)</td>
<td>-0.05</td>
</tr>
<tr>
<td>Bodily Pain</td>
<td>85.90% (18.23%)</td>
<td>83.57% (22.55%)</td>
<td>-2.33% (25.70%)</td>
<td>-0.09</td>
</tr>
<tr>
<td>General Health</td>
<td>71.50% (16.65%)</td>
<td>71.50% (19.57%)</td>
<td>0.00% (17.55%)</td>
<td>0.00</td>
</tr>
<tr>
<td>Vitality</td>
<td>62.17% (19.81%)</td>
<td>65.50% (23.17%)</td>
<td>3.33% (24.08%)</td>
<td>0.14</td>
</tr>
<tr>
<td>Social Functioning</td>
<td>82.93% (23.54%)</td>
<td>82.03% (25.54%)</td>
<td>-0.90% (29.52%)</td>
<td>-0.03</td>
</tr>
<tr>
<td>Role-Emotional</td>
<td>80.00% (29.88%)</td>
<td>74.27% (37.91%)</td>
<td>-5.73% (45.79%)</td>
<td>-0.13</td>
</tr>
<tr>
<td>Mental Health</td>
<td>69.23% (21.22%)</td>
<td>72.80% (20.16%)</td>
<td>3.57% (23.63%)</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Figure 35  Responsiveness testing: comparison of mean OQLQ domain scores at all three times points.

NB: In all cases, 0 is the best quality of life and increasing scores indicate poorer quality of life. Bonferroni corrections applied.

**p < 0.01
***p < 0.001

Scoring
Social aspects is scored from: 0-32
Dentofacial aesthetics is scored from: 0-20
Function is scored from: 0-20
Awareness of dentofacial aesthetics is scored from: 0-16
Table 57  SF-36 data (95% CI) for the responsiveness testing group.

<table>
<thead>
<tr>
<th></th>
<th>PF (PHYSICAL FUNCTIONING)</th>
<th>RP (ROLE-PHYSICAL)</th>
<th>BP (BODILY PAIN)</th>
<th>GH (GENERAL HEALTH)</th>
<th>VT (VITALITY)</th>
<th>SF (SOCIAL FUNCTIONING)</th>
<th>RE (ROLE-EMOTIONAL)</th>
<th>MH (MENTAL HEALTH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire 2</td>
<td>95.00%</td>
<td>92.50%</td>
<td>85.90%</td>
<td>71.50%</td>
<td>62.17%</td>
<td>82.93%</td>
<td>80.00%</td>
<td>69.23%</td>
</tr>
<tr>
<td>(T2)</td>
<td>(90.17% to 99.83%)</td>
<td>(84.69% to 100.0%)</td>
<td>(79.09% to 92.71%)</td>
<td>(62.28% to 77.72%)</td>
<td>(54.77% to 69.57%)</td>
<td>(74.14% to 91.72%)</td>
<td>(68.84% to 91.16%)</td>
<td>(61.31% to 77.15%)</td>
</tr>
<tr>
<td>Questionnaire 3</td>
<td>96.00%</td>
<td>90.83%</td>
<td>83.57%</td>
<td>71.50%</td>
<td>65.50%</td>
<td>82.03%</td>
<td>74.27%</td>
<td>72.80%</td>
</tr>
<tr>
<td>(T3)</td>
<td>(93.68% to 98.32%)</td>
<td>(80.88% to 100.0%)</td>
<td>(75.15% to 91.99%)</td>
<td>(64.19% to 78.81%)</td>
<td>(56.85% to 74.15%)</td>
<td>(72.49% to 91.57%)</td>
<td>(60.11% to 88.43%)</td>
<td>(65.27% to 80.33%)</td>
</tr>
</tbody>
</table>
Table 58  Comparison of SF-36 data with data from the Oxford study (Jenkinson et al., 1993): mean values (SD).

<table>
<thead>
<tr>
<th></th>
<th>PF</th>
<th>RP</th>
<th>BP</th>
<th>GH</th>
<th>VT</th>
<th>SF</th>
<th>RE</th>
<th>MH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(PHYSICAL FUNCTIONING)</td>
<td>(ROLE-PHYSICAL)</td>
<td>(BODILY PAIN)</td>
<td>(GENERAL HEALTH)</td>
<td>(VITALITY)</td>
<td>(SOCIAL FUNCTIONING)</td>
<td>(ROLE-EMOTIONAL)</td>
<td>(MENTAL HEALTH)</td>
</tr>
<tr>
<td>Questionnaire 2 (T2)</td>
<td>95.0% (12.9%)</td>
<td>92.5% (20.9%)</td>
<td>85.9% (18.2%)</td>
<td>71.5% (16.7%)</td>
<td>62.2% (19.8%)</td>
<td>82.9% (23.5%)</td>
<td>80.0% (29.9%)</td>
<td>69.2% (21.2%)</td>
</tr>
<tr>
<td>Questionnaire 3 (T3)</td>
<td>96.0% (6.2%)</td>
<td>90.8% (26.7%)</td>
<td>83.6% (22.6%)</td>
<td>71.5% (19.6%)</td>
<td>65.5% (23.2%)</td>
<td>82.0% (25.5%)</td>
<td>74.3% (37.9%)</td>
<td>72.8% (20.2%)</td>
</tr>
<tr>
<td>Oxford data (Age group 20-24)</td>
<td>91.6% (16.0%)</td>
<td>90.4% (23.7%)</td>
<td>84.3% (19.6%)</td>
<td>74.5% (18.9)</td>
<td>62.6% (18.6%)</td>
<td>87.8% (18.5%)</td>
<td>80.4% (32.7%)</td>
<td>72.0% (16.8%)</td>
</tr>
<tr>
<td>Oxford data (Age group 25-29)</td>
<td>94.0% (12.7%)</td>
<td>89.4% (26.0%)</td>
<td>84.9% (19.6%)</td>
<td>77.8% (17.4%)</td>
<td>61.9% (18.4%)</td>
<td>89.3% (17.3%)</td>
<td>83.7% (30.9%)</td>
<td>73.9% (16.4%)</td>
</tr>
</tbody>
</table>
Figure 36  SF-36 data at T2 and T3 for the responsiveness testing group.

Figure 37  SF-36 data compared with data from the Oxford study 20-24 years age group (Jenkinson et al., 1993).
5.0 DISCUSSION

5.1 GENERAL DISCUSSION

Patient derived indicators are used widely in medicine, particularly in areas such as oncology. It is likely that there will be a role for such indicators in resource allocation in the near future and dentistry cannot afford to fall behind progress being made in this field (Corson et al., 1999). With millions of pounds spent every year on orthognathic treatment, we must know whether patients are benefiting and be able to provide appropriate evidence. By using combinations of generic and condition-specific instruments to research this area, oral health can be compared alongside other diverse medical conditions.

In an ideal world, time and money would be available to undertake “stand alone” clinical trials for quality of life research. However, this is not always the case and sometimes validation studies must be undertaken in conjunction with existing trials (Juniper et al., 1996). If the preliminary work is already established, it is possible to do the validation study in conjunction with another clinical trial.

This study looked at a relatively new area for the study of HRQL. It showed that it was possible to produce a relatively short and simple list of items relating to physical and emotional function and that the scores produced could be statistically analysed. Analysis produced encouraging results and at this stage it is felt that the measure has a number of potential uses in clinical practice and research. With increased emphasis on clinical governance and quality assurance, such a measure should produce important data. It may also be used in conjunction with economic evaluation to provide a basis for efficient allocation of resources in the future.
5.2 DEVELOPMENT OF THE INSTRUMENT

5.2.1 Subjects and sample size

The subjects for both development and testing of the OQLQ were a representative sample of patients treated within the department with regard to age and gender ratio. A small number of psychometricians would argue that it is important that questionnaire items are not influenced by gender, although this argument has only very limited acceptance to date (Garratt, Personal communication). There were some statistically significant differences between males and females, as shown in Table 48, but these were not the subject of interest for the present study.

The study used the following numbers of subjects:

- Item generation 20 patients (plus 16 clinicians)
- Item reduction 44 patients
- Principal component analysis 85 patients
- Validity testing 62 patients
- Responsiveness testing 30 patients

Although a relatively small number of patients were involved in the item reduction phase and in the testing of the instrument, these sample sizes were comparable with those in the development of a number of other condition-specific quality of life measures. Juniper and Guyatt (1991) used 85 patients for the initial testing of their instrument for rhinoconjunctivitis but did not undertake a PCA. Hyland et al. (1994) used 89 patients for the PCA analysis to develop a condition-specific questionnaire for those with breathing disorders. Veldhuyzen van Zanten et al. (1993) in their gastritis study used 24 patients for the item reduction phase and then evaluated the instrument on 55 patients before and one month after treatment. It must also be considered that before and after in their investigation were only a month apart whereas in this study, the longitudinal nature makes it much more difficult to recruit a large sample size.
Larger sample sizes are ideal in PCA to ensure that the groupings are not simply the effects of sampling error (Nunnally and Bernstein, 1994). However, the PCA results in this study gave clinically meaningful information which suggests an adequate sample size.

5.2.2 Item derivation and reduction

The methodology used in the development of this instrument obtained items from several sources, including knowledge of the literature and unstructured interviews with clinicians and patients. This was followed by item reduction, where the items which were chosen most frequently were included in the final measure, and finally exploratory PCA to determine dimensionality of the instrument. Data from the PCA showed that the cumulative proportion of the variance explained by the first four domains was 72.7%, with the first domain alone (social aspects) accounting for 48.8% of the variance (Table 45). Although the variance due to components 3 and 4 was low, their eigenvalues were still above 1.0. When these domains were looked at individually, they were given the following descriptive labels based on the main areas covered in the items:

1st domain: Social aspects;
2nd domain: Dentofacial aesthetics;
3rd domain: Oral function;
4th domain: Awareness of dentofacial aesthetics.

Most Likert type items used in HRQL offer between 4 and 7 responses (Revicki and Leidy, 1998) hence the choice of responses. An even number of responses (1-4) were selected in an attempt to avoid respondents choosing a mid value out of convenience. It was also felt that a larger number of responses would confuse the issue. For example, Cronin et al. (1998) and Wong et al. (1998) used a 7-point scale but it is
then very difficult for respondents to choose between "a moderate problem" and "some problem" or between "a little problem" and "hardly any problem".

Table 47 shows the data for T1 and also includes dimension scores. Not surprisingly, the highest dimension score was for the dentofacial aesthetics domain, followed by social aspects, awareness of dentofacial aesthetics, and finally oral function.

5.3 RELIABILITY TESTING
The results of the internal consistency analysis are shown in Table 49. A high item-total correlation indicates that the item is closely related to the items in the remainder of the domain. It is recommended that values approaching 0.5 should be achieved for the item to remain. Only item 1 and item 5 gave values below 0.5, but the values were thought to be sufficiently close to 0.5 for the items to remain. In addition, item 1 was felt to be fundamental to the instrument and it was somewhat of a surprise that it gave a low correlation. All other items gave values between 0.52 and 0.88. The alpha coefficients for the four components were also high (between 0.83 and 0.93) confirming good internal consistency.

The test-retest analysis compared the two sets of scores for each component using methodology described earlier (Table 50). Repeated administration of the questionnaire produced highly consistent results suggesting stability of the instrument. Intraclass Correlation Coefficients varied between 0.74 for the oral function domain and 0.88 for the social aspects domain. Values above 0.70 are generally considered acceptable.
5.4 VALIDITY TESTING

For purposes of comparing the OQLQ with the VAS and SF-36 a number of hypotheses were generated, which on the whole were confirmed by the data (Tables 51 to 53). All correlations were in the direction hypothesised. All four OQLQ domain scores showed a significant moderate level of correlation with the VAS. The largest correlations were for the domains of social aspects and dentofacial aesthetics (0.45 and 0.44 respectively). It is difficult to predict the relationship between a global question of aesthetics and function, such as the VAS, and the broader patient-assessed instrument. However, the constructs were sufficiently related to expect a moderate correlation and this was confirmed.

The social aspects domain showed a moderate negative correlation with the SF-36 MCS (-0.50) and the scales of Mental Health (-0.47), Vitality (-0.40) and Social Functioning (-0.38). A smaller level of correlation was found with the scale of General Health perception (-0.26). These correlations were all statistically significant. The dentofacial aesthetics domain showed significant correlations of a similar magnitude with the same SF-36 scales. However, this scale was found to have larger correlations with the scales of Social Functioning (-0.48) and General Health perceptions (-0.40) and lower correlations with the scales of Vitality (-0.34) and the MCS (-0.43). Neither of these domains showed a significant correlation with Role-Emotional and the correlation was lower than hypothesised (-0.22 for both domains). However, the correlation was higher than for the remaining three SF-36 components relating to physical health (PF, RP and BP).

The oral function domain showed very little correlation with any of the SF-36 scales or summary scales, the only exception being a small level of correlation with the General Health perception scale (-0.33). This correlation may suggest that eating problems and functional difficulties lead to negative thoughts about general health. There was virtually no correlation with the SF-36 scale of Social Functioning (-0.11). In hindsight, it was felt that the lack of correlation indicated incorrect formulation of
the hypothesis in view of the fact that only one item ("I don’t like eating in public") out of five related to social aspects.

The awareness of dentofacial aesthetics domain showed a small level of correlation with the Mental Health scale (-0.29) and General Health perceptions (-0.26), but only the former was significant and was also at a poorer level than had been anticipated. The correlations with the SF-36 PCS and MCS were both statistically significant (0.32 and -0.35 respectively).

The final hypotheses were confirmed with poor levels of correlation between the OQLQ domains and the SF-36 scales with a primary physical component. The only exception was a moderate level of correlation between awareness of dentofacial aesthetics and the SF-36 PCS (0.32).

Of the twenty moderate negative correlations which were hypothesised, thirteen of the twenty hypotheses were confirmed, although with two borderline results (Table 52). All but four of the twenty hypotheses gave significant correlations even though some were at a small rather than moderate level. Of the sixteen poor levels of correlation which were hypothesised all were supported by the data, although one was at borderline levels (Table 50).

These findings offer support for the validity of the condition-specific OQLQ. The correlations between the OQLQ domains and the PCS and MCS further supported validity of the OQLQ, by showing stronger significant correlations between the OQLQ and the SF-36 MCS. These results provide good evidence for the construct validity of the OQLQ in which the proposed underlying hypothetical constructs were largely supported by the data (Streiner and Norman, 1995). A further type of validity, content validity, is established where items are representative of the area of interest and cover all relevant aspects and where the instrument “makes sense” to the respondent. The OQLQ ensured content validity in that it was based on individual patient experiences of the effects of dentofacial deformity on everyday life.
5.5 RESPONSIVENESS TESTING

Assessment of responsiveness, used data for the first 30 patients who had completed all sub sections of the questionnaires at T1, T2 and T3 (Table 54). Again, sample size was in accordance with the work of Guyatt et al. (1987) who used 13 patients in one study and 28 in another. None of the domains showed a significant change between T2/T1 but the domain which came closest to significance was that of oral function (p=0.06). This p-value lends support to the responsiveness of the oral function domain to orthodontic treatment. Also of note is that the mean values for all four of the OQLQ domains increased during the T2/T1 period, thus showing reduced quality of life. In contrast, those for the VAS decreased from 6.15 to 5.86, showing a positive effect. It appears that, in the short term, there is a decrease in health which is detected by the OQLQ but not by the VAS. The OQLQ thus appears more responsive than the VAS during the T2/T1 period.

There were no significant changes between T2 and T3 for any of the SF-36 scales or for the VAS between T1 and T2. However, between T2 and T3, the VAS showed a significant improvement (p<0.001) as did three of the four OQLQ domains: social aspects (p<0.01), dentofacial aesthetics (p<0.001) and oral function (p<0.001). This result was anticipated, primarily due to the surgery but removal of the orthodontic appliances may also have played a part. The fourth domain, awareness of dentofacial aesthetics, did not alter significantly between T2 and T3. This suggests that respondents still tend to focus on their own and other people’s face/teeth and that this still bothers them to some extent. This is perhaps not surprising after undergoing such major treatment. This is the domain which appears to be least affected by the intervention and this is supported by the magnitude of change in the domain scores. It would be interesting to determine if respondents eventually show significant improvement in this domain when they have adjusted to their new appearance. It is intended to administer the questionnaire again one year following completion of treatment to investigate this.
Standardised Response Means (SRMs) are given in Tables 55 and 56. As anticipated, both the OQLQ and the VAS showed relatively small changes between T1 and T2 but larger levels of change between T2 and T3. The awareness of dentofacial aesthetics domain showed a smaller SRM between T2 and T3 than did the other three domains. The responsiveness of the OQLQ instrument to intervention was therefore found to be very good for three of the four domains with the fourth domain showing less responsiveness.

The VAS was found to have a higher SRM and was more responsive than the OQLQ domains between T3/T2. However, it must be noted that the OQLQ was more responsive to the adverse changes occurring in the T2/T1 period, as represented by the increases in mean scores. It is, therefore, recommended that the VAS be used alongside the OQLQ in future studies. Single item measures are generally less reliable and the ability of the VAS to capture all patient concerns is borne out by the moderate levels of correlation with the OQLQ domains.

The health transition question, which was also included at T3, also supported the responsiveness of the instrument. Twenty-four of the 30 respondents said they felt their dental/facial problems were much better, 5 said better and only one said worse.

5.6 SHORT-FORM 36 DATA

Tables 57 and 58 and Figures 36 and 37 show the data from the Short Form-36 health survey. There were no significant differences for any of the subscales between T2 and T3. Figure 37 also illustrates the similarity between data for these patients and normative data for subjects of approximately the same age. These findings, in conjunction with the SRMs shown in Table 56 illustrate the lack of responsiveness of the SF-36 to the intervention in question. There are serious limitations in the use of generic health measures for this group of patients and this further reinforces
suggestions made by Ware (1993) and Garratt et al. (1996a, 1996b) that ideally both a generic and condition-specific measure should be used in research and/or clinical trials.

Interestingly, Figure 36 shows a slight reduction, although not significant, in the scores for Role-Emotional and Social Functioning. It had been anticipated that both of these scores may increase following treatment. Small increases, although again not significant, were found for Vitality and Mental Health. It is presumably due to the general nature of the questions in the SF-36 that there is insufficient responsiveness to detect the very specific changes occurring after orthognathic treatment.

5.7 GENERAL CONCLUSIONS

Table 59 on the following page lists the necessary criteria for an instrument to be classified as valid, reliable and responsive and looks at how well the OQLQ complies. Although, the OQLQ does not perform perfectly in all areas, initial research suggests that the validity, reliability and responsiveness are sufficiently good to proceed with further investigations using the questionnaire. The processes of investigating psychometric properties of a questionnaire are on-going and not a “once only” process, therefore, further work is required in this area.

This study did, however, show that the OQLQ has good evidence for reliability, validity and responsiveness. The OQLQ was found to be acceptable to respondents and this was supported by the high rates of completion. It is also brief, which makes it suitable for use alongside other instruments, for example, generic measures or other psychometric instruments. It has long been suggested that dentofacial deformity affects quality of life and that orthognathic treatment leads to improved quality of life, but to date these claims were largely unsubstantiated. This study goes some way to supporting these suggestions. The study also highlights the limitations of generic health measures such as the SF-36.
In developing condition-specific measures HRQL, it has been stated that the following five criteria should be fulfilled and it is reassuring to note that all five of these apply to the OQLQ (Guyatt et al., 1989).

1. summary scores should be amenable to statistical analysis;
2. repeated administration in stable patients must yield similar results;
3. when even a small clinically important change has occurred, the questionnaire score should reflect it;
4. the questionnaire should be valid; and
5. the questionnaire should be relatively short and simple.
### Table 59  Validity, reliability and responsiveness: How does the OQLQ compare? (from Jenkinson et al., 1999)

<table>
<thead>
<tr>
<th>REQUIREMENTS</th>
<th>DEFINITION</th>
<th>METHOD OF ASSESSMENT</th>
<th>DOES “OQLQ” COMPLY?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face validity</td>
<td>Do the questions make sense and do they appear to be relevant for the population from which the patients are drawn?</td>
<td>Experts in the field and patients with the condition should be asked to read the questions and assess them in terms of comprehension and relevance.</td>
<td>YES</td>
</tr>
<tr>
<td>Content validity</td>
<td>Is the choice of, and relative importance given to each question appropriate for the phenomenon being measured?</td>
<td>Experts in the field and patients with the condition should be asked to read the questions and assess them in terms of ease of comprehension and relevance and, if weighted, they should also indicate whether the weights appear to appropriately reflect their severity.</td>
<td>YES</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>Does the measure produce results which correspond with a superior measure or does it predict some future criterion value?</td>
<td>Results from one questionnaire may be compared with those of another, but rarely does a “gold standard” exist, except in cases where results from a short form of the questionnaire can be compared with results of an original longer from.</td>
<td>No “gold standard” exists</td>
</tr>
<tr>
<td>Construct validity</td>
<td>Do the results obtained confirm expected relationships or hypotheses?</td>
<td>Results from the questionnaire should be analysed to determine whether it can differentiate between subgroups on which one would expect it to be able to differentiate (i.e. levels of severity).</td>
<td>YES</td>
</tr>
<tr>
<td>Test-retest reliability</td>
<td>Does the measure produce the same results on different occasions when administered to patients who have experienced no change?</td>
<td>Results from the questionnaire will be assessed to determine whether they are the same (or very similar) between administrations.</td>
<td>YES</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>Do the questions in the instrument assess the same underlying phenomenon?</td>
<td>Cronbach’s Alpha is used to determine internal consistency.</td>
<td>YES</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Is the measure sensitive to changes in health status?</td>
<td>Statistical tests used to establish whether a measure is capable of detecting changes.</td>
<td>YES</td>
</tr>
</tbody>
</table>
6.0 CONCLUSIONS

- The OQLQ has good evidence for validity, reliability and responsiveness. It was also found to be acceptable to respondents.
- It would appear that dentofacial deformity adversely affects quality of life and that orthognathic treatment leads to improved quality of life.
- There are serious limitations in the use of generic health measures such as the SF-36 for this group of patients.
- The OQLQ is likely to be of clinical relevance in clinical trials and also in audit.
7.0 SUGGESTIONS FOR FURTHER STUDY

In addition to undertaking further work on the validity and responsiveness of the OQLQ, it is now intended to use the OQLQ in larger scale clinical investigations.

Possible areas of research where the instrument may prove useful include:

- To determine quality of life changes for different malocclusions.
  For example, does treatment of a class III malocclusion produce more improvement in HRQL than treatment of class II individuals?

- To study different demographic groups.
  For example, do younger patients show greater changes than older patients?

- To compare outcomes of single jaw and bimaxillary surgery.
  Does bimaxillary surgery produce greater improvements in quality of life than single jaw surgery?

- To compare different treatment methods.
  For example, does treatment by osteogenesis distraction produce greater quality of life improvements than conventional orthognathic surgery? Is there any difference in quality of life changes in patients treated with or without surgical wafers or in those patients managed with rigid fixation compared with intermaxillary fixation.

The prospect of developing a shorter form of the questionnaire by reducing the number of items could also be investigated.
CHAPTER 4

QUESTIONNAIRE BASED STUDY OF
CHANGES IN PSYCHOMETRIC
VARIABLES IN ORTHOGNATHIC
PATIENTS
CHAPTER 4 REVIEW OF THE LITERATURE

1.0 REVIEW OF THE LITERATURE

1.1 INTRODUCTION

Any survey has the same basic aim, to seek an understanding of what causes some phenomenon (de Vaus, 1996). Techniques for data collection may vary from observation of individual cases to structured or in-depth interviews. However, one of the most commonly used methods is the patient self-completion questionnaire.

There are many criticisms of this approach to research and, although some are justified, many are made by individuals who do not have a thorough understanding of the instruments which are available and the methods of data analysis which should be used. These criticisms may be divided into three groups but are largely unfounded if the data are analysed and used appropriately (de Vaus, 1996):

- Philosophical criticism: For example, “surveys just look at particular aspects of people’s beliefs and actions without looking at the context in which they are occurring” or “some things are not measurable, especially by surveys”.

- Technique-based criticism: Some claim that “surveys are too restricted because they rely on structured questionnaires” or that “surveys are too statistical and reduce interesting questions to totally incomprehensible numbers”.

- Political criticism: The most frequent criticism in this group is that survey research is intrinsically manipulative.

1.2 EXPERIMENTAL DESIGN

The experimental design is one of the major aspects of any survey. Seven basic types of design have been described (de Vaus, 1996):

The classic experimental design

This method involves collecting data at two points in time (usually before and after an intervention). If the experimental group changes significantly more than a control
group, it would normally be concluded that this was due to the intervention. The major problem is in the recruitment of the control group and in retaining contact with the group. It is also dependent on the fact that the two groups are the same to begin with and undergo the same experiences over the experimental period (apart from the intervention). This is almost impossible to ensure.

**Panel design**
This method looks at the same group over a period of time. No definite conclusions can be made regarding changes as a result of the intervention, they are only assumptions and must be tested further.

**Quasi-panel design**
This method is similar to the panel design but has the disadvantage that two different groups are being studied at two separate times. However, it is used in some studies as it avoids problems in keeping track of subjects or may be used where time constraints are a problem.

**Retrospective panel design**
This method obtains information at one time point only and asks the patient to recall information. Although it does not have the problem of keeping track of subjects, there is a major problem of selective memory.

**Retrospective experimental design**
This method has the disadvantages of being retrospective but does attempt to deal with the control group problem.

**Cross-sectional design**
The cross-sectional design survey collects measures from at least two groups but only at any one point in time. The groups must be matched as closely as possible, but this is always a source of error.
One-shot case study design
This is the most primitive type of design because it collects information from only one
group at one point in time. It requires assumptions to be made in order to allow any
causal analysis.

Obviously, the selection of the appropriate method depends on the aims of the study,
the time available and the feasibility of finding a control group.

1.2.1 Selection of instruments for use in questionnaire surveys

Selection of an instrument for clinical research purposes requires careful thought.
Such instruments may take the form of self-completion questionnaires, interviews or
observations. Standardised instruments have three essential criteria (Meyer et al.,
1995):

- Standardised procedures for administration, examiner instructions and scoring.
- Objective scoring criteria.
- A specified preferred frame of reference to guide interpretation of scores. This may
  be norm referenced (how an individual compares with others), criterion referenced
  (to a predetermined gold standard) or non-referenced (identifying the strategies
  that examiners use to solve problems or generate answers).

Meyer et al. (1995) discussed a number of factors which must be considered in the
selection of the preferred instrument:

- Initial considerations.
  What is the purpose of the study? What is the intended sample? What is to be
  measured? How widely used is the instrument? Is the manual well-written and
  comprehensive?
- Psychometric considerations.
  Are reliability, validity and responsiveness acceptable?
• Administration, scoring and interpretability.

Are administration directions clear? Are scoring directions clear? Is computerised scoring available? Is score interpretation clear and compatible with the research question and clinical practice? Are guidelines offered for clinical feedback?

• Practical considerations.

What are the costs and benefits of use? Does the instrument make an independent contribution to the assessment or outcome battery? Are adequate staff, financial and equipment resources available to use the instrument properly? What level of staff expertise and training are necessary?

Only when these questions have been considered carefully should the appropriate instrument be chosen. The increasing number of tests available for use in clinical research makes these particularly difficult but important considerations.

St Leger et al. (1991) described three broad types of questions: factual, for example, age and gender, those dealing with opinion and those dealing with motivation. This can be further divided into open and closed questions (St Leger et al., 1991). Open questions have the advantage that the response depends on the respondent’s own views whereas the closed type uses responses which are of interest to the researcher not the respondent. One way round this is to add a catch-all option such as “other reasons - please state”. Closed or forced-choice questions are easier to deal with analytically but open questions potentially provide more information. The ideal is to pilot a questionnaire with open questions and then devise the definitive questionnaire using the responses given.

St Leger et al. (1991) and de Vaus (1996) also stressed the importance of: using simple language; avoiding very lengthy questions; avoiding ambiguous, very general or leading questions; establishing whether the subjects have sufficient knowledge to answer the questions; using a font size which is large enough to read easily; allowing enough space between questions for responses; printing the questionnaire in a professional style and taking care with questions which may produce a response set.
Response sets are a difficult issue. They occur when a subject agrees with a statement regardless of what they really believe (an acquiescent response set) or when the patient answers in such a way as to make themselves look good (a social desirability response set).

### 1.2.2 Question Formats

There are a number of different response formats when asking closed questions (de Vaus, 1996):

- **Likert-style format**
  A question is presented and the respondent is then asked how strongly they agree or disagree. It can have the problem of creating an acquiescent response set problem but is still a very popular format particularly in psychology questionnaires.
  
  The response format generally takes the form:
  
  - Strongly agree
  - Agree
  - No feelings one way or the other
  - Disagree
  - Strongly disagree

- **Semantic differential format**
  How would you describe yourself?

    Shy 1 2 3 4 5 6 7 Outgoing

- **Checklists**
  This approach lists a series of responses and asks the respondent to mark those which apply to them.
• Ranking
Questions of this format give a number of alternative answers and ask people to rank them, for example, from the most desirable to the least desirable.

• Attitude choices
A number of different responses are provided and the respondent is asked to select the view closest to their own.

1.2.3 Administering questionnaires

Questionnaires may be completed by the respondent or by the interviewer. There are three main methods of administration, all of which have advantages and disadvantages:
• Face-to-face interviews;
• Telephone surveys;
• Mail questionnaires.

Dillman (1978, cited in de Vaus, 1996) found response rates of 85% in telephone surveys and 60-75% for mail surveys of the general public. This percentage further increased if a specific group was being surveyed and the questions were relevant to them. Dillman argued that although face-to-face interviews have always claimed to be superior, this is not necessarily the case. The method of administration does affect which type of questions can be asked. Mail questionnaires are not suitable for large numbers of open-ended questions and less complex questions must be asked than in a face-to-face situation. However, mail surveys may be better at obtaining accurate answers as there is less chance of social desirability bias. The optimum length of mail questionnaire should be considered carefully and Dillman (1978, cited in de Vaus, 1996) found that 125 items or 12 pages was the optimum for general public surveys but longer questionnaires were acceptable for specific groups where the questions were of particular relevance.
The funding available may also have a major impact on the decision regarding which method to use. Large studies involving face-to-face interviews incur considerably greater costs than telephone interviews or mail surveys. However, for a small survey on a local level, the amount of variation is likely to be less.

1.3 CHOICE OF PSYCHOMETRIC OUTCOME MEASURES

It is not unreasonable to anticipate that disorders such as anxiety and depression may alter during interventions or may affect satisfaction with outcomes. For this reason, they are commonly measured in psychosocial outcome studies.

1.3.1 Anxiety

State-Trait Anxiety Inventory (Spielberger et al., 1983)

Although there are a number of instruments which measure anxiety, this is probably the most widely used in psychological and clinical research. Anxiety is generally viewed as taking two forms: the state form (transitory feelings of fear/worry) and the trait form (the stable tendency to respond anxiously to stressful situations). The State-Trait Anxiety Inventory (STAI) was developed during the 1960s and revised in 1983 (Spielberger et al., 1983). The questionnaire was developed from a large pool of statements and was extensively tested on college students. The instrument consists of 40 items, 20 of which measure state and 20 measure trait anxiety. Each item is scored on a 4-point scale. It takes approximately 10 minutes to complete and produces scores between 20 and 80 for both sections.

The alpha coefficient has been reported to be high for both state ($r=0.93$) and trait ($r=0.90$) anxiety, which indicates good internal consistency. In addition, test-retest scores showed stability for trait anxiety (with values ranging from 0.65 to 0.86) although this was less for the state scale (0.16 to 0.62). This latter finding is hardly surprising as this is a measure of transient anxiety. Validity testing was undertaken by
comparing the scale with other anxiety scales and this produced correlations of between 0.52 and 0.80. In support of the construct validity, the state component showed higher mean values in stressful situations (Spielberger et al., 1983).

1.3.2 Depression

A wide number of instruments are now available for the assessment of depression. The majority will not make a diagnosis of depression but rely on the provisional diagnosis having been reached already. Some of the scales measure only depression, others measure a combination of depression and anxiety.

Beck Depression Inventory (Beck et al., 1961)
The Beck Depression Inventory (BDI) covers a wide range of items from sadness and sense of failure to self-dislike, social withdrawal and indecisiveness. There are 21 items with scores ranging from 0 to 3 and the total score may, therefore range from 0 to 63. A short-form with 13 items is also available. The 21 item scale takes 5-10 minutes to complete and although the authors did not give cut-off points, scores can be readily analysed as continuous data. However, the usual scoring guides based on normative data suggest 0-9 = normal; 10-15 = mild; 16-19 = mild/ moderate; 20-29 = moderate/ severe and >29 = severe depression.

The BDI has been shown to have good psychometric properties and to be sensitive to change. Beck et al. (1988) showed that the concurrent validity of the BDI was high. Mean correlations with the Hamilton Scale and clinical ratings were over 0.70 for psychiatric patients and over 0.74 for non-psychiatric patients. It was also found to be sensitive to types of depression and to distinguish depression from anxiety. A study of medical in-patients showed a correlation of 0.75 between the BDI and Hamilton Scale (Schwab et al., 1967). Test-retest correlations at 2 to 5 weeks have been reported to be in the region of 0.90 (Beck et al., 1961).
Although, some authors have levelled criticism at the BDI, many experienced researchers have used this scale to produce very useful data. It has been argued that the BDI is so well established as to require serious consideration for any study requiring self-evaluation of depression. A number of other instruments measuring depression also exist, for example the Hospital Anxiety and Depression Scale (Zigmond and Snaith, 1983), Goldberg’s General Health Questionnaire (Goldberg and Williams, 1988) and Zung’s Self-Rating Depression Scale (Zung et al., 1965).

1.3.3 Self esteem

The concept of self esteem, or perceived self-worth, is part of the wider construct of self-concept. It is a personal resource which may moderate the effects of conditions or events. Self esteem is a dimension of life satisfaction and is an important part of general assessment of life (Andrews and Withey, 1976). It is developed and maintained through personal interactions and is believed to become increasingly important with age.

Rosenberg Self Esteem Scale (Rosenberg, 1965)

The Rosenberg Self Esteem (RSE) Scale is the most widely used measure of self esteem in health psychology. It was developed by Rosenberg (1965) for a study of students in New York but has also been shown to be suitable for older subjects (Kaplan and Porkorny, 1969). In the original study, self esteem scores were correlated with characteristics such as leadership and participation in college activities. It is a self administered 10 item scale with a four point scale of agreement and takes approximately five minutes to complete. Items are scored from 1 to 4 in the direction of negative self esteem with low scores indicating high self esteem. Half the items are expressions of positive self esteem and half are negative.
Rosenberg (1965) chose items which he felt had face validity and there is some evidence of the construct validity of the scale. Robinson and Shaver (1973, cited in Bowling, 1997) reported that the scale correlated well \((r=0.59-0.60)\) with the Coopersmith’s Self Esteem Inventory (Coopersmith, 1981), although this did depend on the scoring method. There are few other measures available with which to compare this scale. Reliability was demonstrated to be good, with reproducibility coefficients of 0.85-0.92 (Rosenberg, 1965). The brevity and simplicity of this scale make it a popular choice. A number of other scales may be used to measure self esteem including the Tennessee Self-Concept Scale (Fitts, 1965, cited in Bowling, 1997) and the Coopersmith Self Esteem Inventory (Coopersmith, 1981).

### 1.3.4 Social support

Social support is a broad term referring to social, emotional and other supports which are provided by an individual’s social contacts. Cohen and Wills (1985) described four different functions of social support:

- **Esteem support**: This is the effect of others in promoting someone’s feelings of self esteem.
- **Information support**: This is the useful and necessary information which is acquired from social contacts.
- **Instrumental support**: The physical support gained from others, for example, money.
- **Social companionship**: This is support derived from spending time with others, for example, during leisure time, activities etc.

Others have distinguished between emotional and practical support (Power et al., 1988) and frequently a distinction is made between perceived (subjectively rated adequacy of support) and actual support (number of individuals). The concept of social support affecting outcome of health intervention is an interesting one. Social support is commonly seen as a resource which can be mobilised in times of crisis to
CHAPTER 4  REVIEW OF THE LITERATURE

protect against health threats. This is the “stress-specific hypothesis”. It suggests that
social support operates selectively under stress and that it is only under these
circumstances that individuals with high levels of social support will fare better than
those with low levels of support. The “universal effect model” (Cohen, 1988)
suggests that high levels of social support will be associated with better health even in
the absence of stress. A third theory is that the absence of social support is itself a
stressor. Thus, social support is a complex phenomenon and its effects can be
mediated in a number of different ways. These effects may be direct, for example
reducing or eliminating stress responses by changing physiological processes, or
indirect, by bringing about changes in health-related behaviour or by facilitating
health-enhancing behaviours (Cohen, 1988).

There are a number of social support measures available but these are reduced if
those which are lengthy and time consuming to complete are excluded. In addition, a
number of instruments designed to measure quality of life also include items on social
support. There is no instrument which measures all components of social network and
support with acceptable levels of validity and reliability.

**Short Form Social Support Questionnaire** (Sarason *et al.*, 1983, 1987)

This 6 item self-administered questionnaire was developed from an original 27 item
questionnaire. It was developed by asking a large group of students to list people who
provide them with support and to list situations where social support might be
important. It produces two measures: availability of support and satisfaction with
support. The authors suggest that these two scores are treated independently. There
are no norms, the scores in any study being relative measures.

It has been found to be a viable measure of support. Sarason *et al.* (1987) found high
levels of internal consistency for both subscales (alpha=0.97 for availability and 0.94
for satisfaction) and high test-retest reliability (0.90 for availability and 0.83 for
satisfaction). Other measures of social support include the Inventory of Socially
Supportive Behaviour (Barrera and Ainlay, 1983) and the Interview Schedule for Social Interaction (Henderson et al., 1980).

1.3.5 Body image

Body image may be a source of distress for individuals of all ages and has become more of an issue in recent years with the large amount of research focusing on individuals with eating disorders (Lautenbacher et al., 1993; Nelson and Gidycz, 1993). The psychological representation of one’s body depends on how the visual and somatosensory stimuli are processed and integrated. Sensory inputs are integrated into the body self in a series of complex steps. Under some circumstances the visual input and the somatosensory inputs are not well integrated into the “body self” (which encompasses various aspects of body image including body size perception), although in everyday life this appears to be largely successful (Lautenbacher et al., 1993).

Body image develops over a period of time and is influenced by four main factors (Belfer et al., 1982):

- **Cognitive function.** The level of cognitive function influences the individual’s perception of his/her body.
- **Perception of body stimuli.** Tactile stimuli can be altered by physical anomalies and may affect body image as a result. An individual with facial disfigurement receives sensory input through social interaction and this may frequently be perceived as negative thus affecting body image.
- **Stimuli from the environment in the form of comparison.** As an individual compares their body with parents, and later peers, they become aware of differences which have an effect.
- **Response from others.** The response of parents, friends and peers will influence an individual’s ability to cope with a deformity.
If body image is accepted as being not simply a function of objective appearance but a complex psychological concept related to the mental representation of self, then a change in body image resulting from surgical procedures must be a complicated process. It is likely that this may be influenced by the surgery itself and the response of others to the surgical result. This reintegration phase may take varying time periods depending on the post-operative recovery and also the patient’s expectations. Belfer et al. (1982), in their study of facially disfigured children undergoing reconstructive surgery, found that the reintegration phase took in the range of six weeks to six months. In addition, body image is also related to self esteem. A body or facial defect may affect self esteem directly, by negative feedback from others, or indirectly, by affecting the ability to master certain developmental tasks. Body image development affects self esteem in both "normal" individuals and those with body defects.

Much of the work on body image has involved adolescents with eating disorders (Lautenbacher et al., 1993; Nelson and Gidycz, 1993) and there are few scales available for the measurement of body image in a group of subjects such as those undergoing orthognathic surgery.

**Secord and Jourard’s Body Cathexis Scale**

This scale was not designed specifically for facial deformities. However, recent studies investigating facial deformities have used a revised version with a specific section asking about facial features (Kiyak et al., 1982a, 1982b, 1984; Finlay et al., 1995). The revised scale is designed for self-completion and has 27 items. The respondent is asked to record their feelings about each body part to produce an overall body image score, a facial image score and also scores about specific areas such as the teeth, chin etc. The five responses are given a score of 1 to 5 and scores are summed to provide a total score. Higher scores represent lower satisfaction with body image. There are few other scales or instruments which measure body image or facial aspects of body image.
1.4 PSYCHOLOGICAL ASPECTS OF ORTHOGNATHIC TREATMENT

Many studies have looked at changes in psychological profile and personality characteristics following orthognathic treatment. Early studies tended to be retrospective (Hutton, 1967; Crowell et al., 1970). More recent studies have been prospective, but the baseline which has been used for pre/post treatment comparisons is typically prior to surgery rather than prior to any active treatment at all (Kiyak et al., 1982a, 1982b, 1984; Flanary et al. 1990; Barbosa et al., 1993; Finlay et al., 1995). There is very little information regarding potential psychological effects of pre-surgical orthodontics.

Studies which have looked at pre- and post-surgery changes include the longitudinal studies undertaken by Kiyak et al. (1982a, 1982b, 1984, 1985, 1986) who found good psychological adaptation following orthognathic treatment. Finlay et al. (1995) reached similar conclusions. Flanary et al. (1990) investigated psychological adjustment and self-concept in orthognathic patients and found them to be relatively healthy and well adjusted both pre- and post-operatively. Significant improvements were noted post-operatively in the following areas: psychoses; neuroses and personality disorders; integration and all sub-scales of self-concept.

Although such studies have produced very important information regarding psychological aspects of treatment, it is difficult to make definite conclusions because there was no baseline measurement before any active treatment commenced.
1.5 MULTILEVEL MODELLING

1.5.1 Introduction to multilevel modelling

One of the newer developments in statistics is multilevel modelling (Goldstein, 1987, 1995) or hierarchical linear modelling. Multilevel modelling (MLM) is a method which should be used where data falls naturally into hierarchical structures. A statistical method which uses this type of structure, rather than ignoring it, is important as failure to recognise hierarchical structures can result in falsely narrowed confidence intervals, hence an inflation of significance (Albandar and Goldstein, 1992).

In medicine and social sciences, multilevel and hierarchically structured data are common (Goldstein et al., 1998). Using the example that children are taught within classes which are taught within schools, children would be assigned to level 1, classes to level 2 and schools to level 3. In medicine, several hospitals may be chosen to test a new drug and individual patients who are given the drug would be studied. Patients would then be at level 1 and the hospitals would be at level 2. In dentistry, periodontal research has made use of multilevel modelling in the past (Albandar, 1990; Axtelius et al., 1999). In this case, disease sites (level 1) are grouped within teeth (level 2), which are grouped within subjects (level 3).

Previously multilevel data would have been dealt with in three ways:

- Ignoring the levels completely which obviously results in erroneous statistical analyses with serious implications on results and conclusions.
- Reducing higher levels to lower levels (atomistic fallacy) results in inefficient generalisations and involves estimating many more coefficients than with multilevel procedures. It provides no information about the variation of the underlying population.
- Reducing lower levels to higher levels (ecological fallacy) results in loss of information and interpretations are very difficult to make. Detailed information may be lost because the focus of interest may lie at a lower level in the data organisation.
The multilevel technique prevents these compromises having to be made (Söderfeldt and Palmqvist, 1998). Multilevel modelling techniques respect complex data structures rather than work around them. For example, multilevel analysis allows longitudinal data to be considered as another level from the hierarchical structure. In addition, data do not need to be balanced, measurements may be available on only one occasion for some data but several occasions for other. Provided this information is missing at random, the technique can deal readily with the problem. Combined with issues such as the problem of within-subject variance being as large or larger than between-subject variation and the difficulties of longitudinal data analysis, then it becomes clear that more sophisticated techniques are required in dental research than have been used to date.

If the effect of several independent variables on one dependent variable is desired, multiple regression analysis is used. Multivariate multiple regression analysis is one of the useful applications of multilevel modelling in dentistry and this may be thought of as multiple regression for several equations performed in parallel, the advantage of which is that more information is used with commensurate increase in statistical power. An underlying assumption of standard single-level regression analysis is that all the observations are independent, failure to recognise hierarchical data and the use of single level analysis violates these basic statistical assumptions (Söderfeldt and Palmqvist, 1998).

Therefore, the application of multilevel modelling has a number of important benefits including:

- The estimation of effects within individual units is improved because the estimate for a regression model for an individual subject is enhanced by “borrowing” strength from similar estimates for other subjects.
- The model allows partitioning of variance and covariance components between the different levels (i.e. “within” and “between” subjects).
- Improved formulation and testing of hypotheses regarding cross-level effects (for example, how a subject characteristic such as oral hygiene status might affect the
relationship between site-level plaque and loss of attachment within periodontal patients).

Figure 38 Example of a multilevel data structure in dentistry.

1.5.2 Multilevel modelling in dentistry

Multilevel modelling has been used in very few areas of dentistry. Work has tended to focus on periodontal data because of the obvious hierarchical nature of the data. Sterne et al. (1988) undertook one of the early studies using a developmental form of multilevel modelling in a clinical trial investigating the use of metronidazole in chronic periodontal disease. A subject-only analysis produced borderline results, with a positive effect only for the most severe cases but multilevel modelling showed an overall significant treatment effect.

Albandar (1990) also used the technique in the field of periodontology and studied predictors for the progression of periodontal disease. The results were in contrast to those obtained from earlier single-level analyses of the same data (Albandar et al., 1987). Data which were previously significant using single-level techniques became non-significant when multilevel modelling was used. Albandar (1990) stressed that
these problems highlighted the importance of using the correct statistical techniques for hierarchical data. Axtelius et al. (1999) studied the progression of periodontal disease in 22 patients. The model included data regarding pocket depth probing but was also extended to include independent variables for sites, teeth and individuals. They concluded that multilevel analysis was the most appropriate statistical technique for these type of data.

Söderfeldt and Palmqvist (1998) used multilevel modelling to study longevity of fixed partial dentures, retainers and abutments based on the hierarchical data structure of “tooth, arch, patient”. They found that conventional logistic regression over- and under-estimated probabilities of survival in different cases when compared with multilevel modelling and concluded that the technique provided a better understanding of the interaction between various factors affecting longevity of restorations.

Some multilevel modelling research has been undertaken in the field of orthodontics and has important applications in the field of longitudinal craniofacial growth. Buschang et al. (1989) investigated longitudinal mandibular growth and Hoeksma and van der Beek (1991) used multilevel modelling of longitudinal cephalometric data to analyse hypothetical growth curves. Hoeksma and van der Beek (1991) showed how both individual and average growth could be modelled. The authors proposed four benefits of using the multilevel modelling technique to study longitudinal growth: the model described individual and average growth curves in a meaningful way; the models were flexible and described growth curves for many different growth patterns; missing data could be handled readily and finally, the technique could be used with many different research study designs.
2.0 AIMS

The aims of this study were to use statistical modelling to:

- determine whether orthognathic patients differed psychologically from a control group of non-patients.
- establish whether any psychological changes occurred during the pre-surgical orthodontic phase and whether it was acceptable to use pre-surgery as a baseline.
- establish whether the psychological profile of patients altered during treatment.
3.0 MATERIALS AND METHODS

3.1 CONTENTS OF THE QUESTIONNAIRE

Ethical approval was received from the Joint Research and Ethics Committees of the units involved. Respondents were provided with an information sheet and advised that all answers would be treated in the strictest confidence. They were also informed that failure to participate in the study or withdrawal during the study would not affect their care. All respondents completed a consent form.

A questionnaire was developed in conjunction with a liaison psychiatrist and professor of psychology to assess various aspects of the respondent’s psychological profile. Areas were selected which it was felt may play a part in the individual responses during treatment. Scales were selected which had been tested rigorously for validity and reliability in their development and in previous studies and had been used widely in health psychology research in the past. The length of the questionnaire complied with advice based on work conducted by Dillman (1978, cited in de Vaus, 1996). Dillman recommended that 12 pages was the optimum length for general public respondents although could be a little longer for specific groups where the questionnaires were of particular relevance.

The questionnaire used a number of psychometric instruments which are described more fully in the review of the literature:

- **Anxiety**: State-Trait Anxiety Inventory (Spielberger et al., 1983).
- **Depression**: Beck Depression Inventory (Beck et al., 1988). One item (related to suicidal ideation) was omitted from this scale due to difficulties experienced by colleagues at the same institute in an earlier study.
- **Perception of social support**: Short Form Social Support Questionnaire (Sarason et al., 1987).
- **Self esteem**: Rosenberg Index of Self Esteem (Rosenberg, 1965).
- **Body image**: Revised version of Secord and Jourard’s Body Cathexis Scale (Secord and Jourard, 1953; Kiyak et al., 1982).
• **Facial body image**: This is a sub-scale of the body image scale. Kiyak and colleagues used items 5-12 (Appendix 6) to represent facial body image although “profile” and “overall facial appearance” were also included in this study (Kiyak, Personal communication).

**Table 60** Scoring of questionnaires (score range in parentheses).

<table>
<thead>
<tr>
<th>Instrument/ Questionnaire</th>
<th>Scoring</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety (20-80)</td>
<td>Increasing anxiety</td>
<td>Higher scores</td>
</tr>
<tr>
<td>Trait anxiety (20-80)</td>
<td>Increasing anxiety</td>
<td>Higher scores</td>
</tr>
<tr>
<td>Depression (0-60)</td>
<td>Increasing depression</td>
<td>Higher scores</td>
</tr>
<tr>
<td>Social support: number (0-9)</td>
<td>Increasing social support</td>
<td>Higher scores</td>
</tr>
<tr>
<td>Social support: satisfaction (0-6)</td>
<td>Increasing social support</td>
<td>Higher scores</td>
</tr>
<tr>
<td>Self esteem (10-40)</td>
<td>Increasing self esteem</td>
<td>Lower scores</td>
</tr>
<tr>
<td>Body image (27-135)</td>
<td>Increasing satisfaction with body image</td>
<td>Lower scores</td>
</tr>
<tr>
<td>Facial body image (10-50)</td>
<td>Increasing satisfaction with facial body image</td>
<td>Lower scores</td>
</tr>
</tbody>
</table>
3.2 RELIABILITY

Although all instruments selected had been subjected to extensive psychometric testing during their development and had been shown to be reliable, it was felt necessary to establish whether the instruments were reliable for this specific group of patients. Twenty five pre-treatment orthognathic patients were asked to complete a questionnaire and return it in a stamped addressed envelope. Six weeks later an identical questionnaire was sent, with a covering letter explaining the reason behind the second questionnaire. Again, a stamped addressed envelope was included. All 25 respondents completed the first questionnaire but only 19 (76% response rate) also returned the retest questionnaire.

3.3 SUBJECTS

3.3.1 Comparison of patients and control group: sample size

Sample size calculation was undertaken using non-parametric options in nQuery Advisor® Release 3.0. The Mann-Whitney programme is based on an estimation of the differences between the two groups. However, as there was little previous data from which to draw useful information, assumptions regarding what magnitude of differences would be important were based on advice from colleagues in the field of psychology and also on clinical judgement. It must be acknowledged that smaller differences were not discounted as unimportant. The programme uses a common standard deviation (SD) from the two groups and SDs were obtained from the first 20 subjects (10 experimental and 10 control) who completed questionnaires. Although this data could also have been used to estimate the clinically relevant differences, it was felt that with only 20 subjects the differences in scores may be non-representative and misleading. This method of sample size calculation assumes that single level analysis is being undertaken and, in practice, because multilevel analysis was used in this study, the sample size quoted is probably over-pessimistic.
Table 61  Sample size calculation.

<table>
<thead>
<tr>
<th></th>
<th>Estimated clinically important difference between groups (common SD)</th>
<th>Required Sample size in each group 0.05 significance; 80% power</th>
<th>Required Sample size in each group 0.05 significance; 90% power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spielberger’s state anxiety (Max score = 80)</td>
<td>4 (8.12)</td>
<td>71</td>
<td>95</td>
</tr>
<tr>
<td>Spielberger’s anxiety trait (Max score = 80)</td>
<td>4 (7.18)</td>
<td>56</td>
<td>75</td>
</tr>
<tr>
<td>Beck’s depression inventory (Max score = 60)</td>
<td>3 (3.21)</td>
<td>22</td>
<td>30</td>
</tr>
<tr>
<td>Secord and Jourard’s body cathexis scale (Max score = 135)</td>
<td>5 (10.36)</td>
<td>73</td>
<td>98</td>
</tr>
<tr>
<td>Secord and Jourard’s body cathexis scale: facial component (Max score = 50)</td>
<td>4 (4.62)</td>
<td>25</td>
<td>34</td>
</tr>
<tr>
<td>Rosenberg’s index of self esteem (Max score = 40)</td>
<td>2 (4.30)</td>
<td>79</td>
<td>106</td>
</tr>
<tr>
<td>Sarason’s social support: number (Max score = 9.0)</td>
<td>1.0 (2.83)</td>
<td>134</td>
<td>179</td>
</tr>
<tr>
<td>Sarason’s social support: satisfaction (Max score = 6.0)</td>
<td>1.0 (1.19)</td>
<td>27</td>
<td>35</td>
</tr>
</tbody>
</table>

NB: In reality numbers required should be lower than this due to statistical method employed.
3.3.2 Comparison of patients and control group: subject 
recruitment and response rate

Patients were recruited from three orthodontic/maxillofacial surgery units. The units were chosen based on their similar approach to orthognathic treatment and, in addition, a number of staff worked at two of the three units involved which provided consistency. However, some differences exist amongst the three units. One unit is a teaching hospital (unit 1) and the other two (units 2 and 3) are district general hospitals in areas with potentially different populations. All patients who started orthognathic treatment in these three units over a 12 month period (June 1997 to May 1998) and who fulfilled the selection criteria already described in Chapter 2 were asked to take part. Only two patients expressed a wish not to be involved.

The questionnaire was distributed to a total of 88 patients, who had decided to proceed with treatment. Patients were asked to complete the questionnaires at home as it took in the region of 30-40 minutes to complete. They were provided with a stamped addressed envelope and asked to return the questionnaire within the following week. Respondents who did not reply within two weeks were contacted by telephone. The response rate for the first questionnaire was 97% (85 responses).

In order to compare these patients with a control group of non-patients, 106 individuals were recruited from local colleges and offices. The control group was matched by age, gender and ethnicity as closely as possible. A total of 106 individuals were approached and 96 questionnaires were returned (response rate 91%).

Data was complete for the Beck Depression Inventory but all other scales had some data missing. One respondent failed to answer for state anxiety; three for trait anxiety; ten failed to complete the social support questions and four omitted the self esteem and body image questions. The non-completion of the social support measure was felt to reflect the additional effort required for its completion.
3.3.3 **Longitudinal study (patients only): sample size**

Sample size calculation was again undertaken using nQuery Advisor® Release 3.0. As in Chapter 2, it was the difference between start of treatment (T1) and end of treatment (T3) values which was of main concern. Assumptions regarding what magnitude of differences would be important were judged as in section 3.3.1 with standard deviations of T3/T1 differences obtained from the first 8 subjects who completed all questionnaires.

**Table 62**  Sample size calculation.

<table>
<thead>
<tr>
<th></th>
<th>Estimated clinically important difference between T1 and T3 (SD from 8 subjects)</th>
<th>Required Sample size in each group 0.05 significance; 80% power</th>
<th>Required Sample size in each group 0.05 significance; 90% power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spielberger's state anxiety (Max score = 80)</td>
<td>5 (12.32)</td>
<td>50</td>
<td>66</td>
</tr>
<tr>
<td>Spielberger's trait anxiety (Max score = 80)</td>
<td>5 (14.44)</td>
<td>68</td>
<td>90</td>
</tr>
<tr>
<td>Beck's depression inventory (Max score = 60)</td>
<td>3 (8.11)</td>
<td>60</td>
<td>79</td>
</tr>
<tr>
<td>Secord and Jourard's body cathexis scale (Max score = 135)</td>
<td>5 (10.39)</td>
<td>36</td>
<td>48</td>
</tr>
<tr>
<td>Secord and Jourard's body cathexis scale: facial component (Max score = 50)</td>
<td>4 (5.94)</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>Rosenberg's index of self esteem (Max score = 40)</td>
<td>3 (4.49)</td>
<td>20</td>
<td>26</td>
</tr>
</tbody>
</table>

NB: In reality numbers required should be lower than this due to statistical method employed.
3.3.4 Longitudinal study (patients only): subject recruitment and response rate

It was estimated that the first 65 respondents to be recruited should complete treatment within the time constraints of this research. Therefore, these 65 individuals were asked to complete questionnaires on three separate occasions:

T1 Prior to any active treatment (following the visit to the Joint Orthodontic/Maxillofacial Surgery Clinic).

T2 Following pre-surgical orthodontics and prior to surgery (at least six weeks prior to surgery).

T3 Four to six weeks post debond.

The response rate at T2 was 95% (62 of 65 questionnaires returned). Two patients did not receive the questionnaire in time to complete it 6 weeks prior to surgery and one was not returned. Unfortunately, during the study one unit moved sites and one experienced severe problems with bed availability. Both factors led to cancellation of a number of operations and only 55 of the 65 respondents had undergone surgery at the time of final data analysis. Of these 55 individuals, 46 were debonded and 42 returned the T3 questionnaire (response rate 91% of completed cases).
3.4 STATISTICAL ANALYSIS

Previous studies which have looked at psychosocial factors in relation to orthognathic treatment have analysed each of the outcome measures (for example, self esteem or body image) separately. By analysing components separately, the data is not being explored fully. In addition, the hierarchical nature of the data has been ignored in the past (i.e. several outcomes occurring within each subject). For these reasons multilevel modelling (MLM) was adopted using the MLwiN statistical software (Rasbash et al., 1999). Data were initially entered on an SPSS spreadsheet and subsequently converted for use by MLwiN.

3.4.1 Multivariate data

The data formed a natural hierarchy with outcome measures occurring at the lowest level of the data structure clustered at the higher level within subjects. Analysing multiple outcomes independently has the advantage that it is possible to interpret the results meaningfully in terms of each outcome. However, the disadvantage is that not all available information would be used at once, thus reducing the statistical power of the analytical process (Cunningham et al., 2000; Gilthorpe and Cunningham, 2000). Multivariate analysis overcomes this by analysing all outcome measures simultaneously. It is more efficient and effective to evaluate all outcomes using multivariate techniques because this optimises the estimation of statistical significance (Goldstein, 1995).

Figure 39 Multilevel data structure.
3.4.2 Multiple regression

Multiple regression analysis was chosen to analyse the study data for the first part of the study because it was not feasible to match every member of the experimental group with a control group respondent with respect to possible confounding factors (age, gender, ethnicity). This would have required paired matching, but it was only feasible to ensure that the overall demographic data were similar for the two groups. Thus, differences in any of the outcomes may be due to demographic differences between patients and controls, and not just intrinsic differences between those with dentofacial deformity (patients) and those without (controls).

Multiple regression analysis was therefore undertaken where patient characteristics were considered simultaneously for their impact on the eight outcome measures. This method allowed any underlying demographic differences to be accounted for whilst evaluating the impact of dentofacial deformity (i.e. whether the patient was in the experimental or control group). The outcome measures were analysed simultaneously using multivariate techniques, adopting the MLM approach.

The same technique was selected for the remainder of the study because, in addition to the intervention (orthognathic treatment), demographic characteristics (age; gender; ethnicity) and other patient characteristics (duration of treatment; unit of treatment; patient-perceived severity of malocclusion; type of malocclusion) may also have influenced the outcome measures.
3.4.3 Data management

Measures of Normality: skewness and kurtosis

It is a requirement of standard regression analysis that model residuals follow a Normal distribution. For models with no covariates (i.e. no explanatory variables), this distributional requirement extends to the outcome variable. Therefore, as a precursor to modelling, each outcome variable was assessed for Normality. Values which were skewed or kurtosed were transformed to produce a more Normal distribution using functions of the natural logarithm. Ideally both skewness (symmetry) and kurtosis (smoothness) should be zero after transformation. However, symmetry is more important than smoothness. For some data sets there was no ideal transformation, in which case the aim was to achieve zero skewness and minimal kurtosis.

For the longitudinal studies, where differences in scores were of interest rather than the raw scores, this step was not required as differences in scores tend to be much more Normally distributed than actual raw scores, and this was confirmed by data inspection.

Variable centring

Further data preparation involved centring all explanatory variables whose range did not contain zero; in this case, age for the first study and age and duration of treatment for subsequent studies. Centring is a process where all values have their mean (or a figure close to the mean) deducted. This process is advisable prior to regression analysis since the inclusion of explanatory variables whose range do not contain zero, increases uncertainty in the modelled coefficients and inflates their associated confidence intervals (Gardner and Altman, 1993). This process therefore provided a meaningful interpretation of the regression coefficient when the adjusted age (or duration) variable adopts the value of zero.
Standardisation

Assessment of psychometric measures is difficult, particularly when the scales are abstract and difficult to quantify. This makes comparison of differences in each scale difficult to interpret across groups. To provide a similar range of values across all outcome measures, each variable was scaled by standardisation (Rice, 1988). This involves dividing by the variable standard deviation (SD). For convenience, each measure is also centred prior to standardisation, giving each outcome a mean of zero and SD of 1. Differences in one outcome can then be directly compared in magnitude with differences in other outcomes.

Standardisation of changes in scores (as in the longitudinal study) involved only dividing by the standard deviation of the changes.

Coding prior to modelling

Gender and dentofacial groups were binary variables and were therefore allocated values of 1 or 0 for: male/female and experimental/control group. Antero-posterior (A-P) skeletal relationship was categorised in the same way for the longitudinal analysis: class II and class III.

The ethnic groups recorded were Caucasian, South Asian (Bangladeshi, Bengali, Indian and Pakistani), black (black African, black Afro-Caribbean and black other) and other (including Moroccan, Cypriot and Persian). Since this variable had more than two categories, “dummy” variables were required to adequately represent the options. Groups were contrasted with the largest group (Caucasian). The reference category, against which others were contrasted, was that for which all dummy variables adopted the value zero. The set of dummies was treated as a single variable and, if the categorical variable attained significance, all dummies remained in the model, even though they may not be significant in isolation.
The unit/hospital in which the respondent was treated and the respondent’s own perceived severity of their problem at the start of treatment (mild, moderate or severe) were also categorised in the same manner and contrasts were made with the largest group (unit 1 and severe, respectively).

3.4.4 The multivariate multilevel statistical analysis

Multivariate regression model

The multivariate multilevel method yields simultaneous regression equations:

\[ y^{(k)} = \beta_0^{(k)} + \beta_1^{(k)} \times \text{facial group} + \beta_2^{(k)} \times \text{gender} + \beta_3^{(k)} \times \text{centred age} + \beta_4^{(k)} \times \text{ethnic1} \]
\[ + \beta_5^{(k)} \times \text{ethnic2} + \beta_6^{(k)} \times \text{ethnic3} + e^{(k)} \]

(Equation 1)

Where: \( k = 1, \ldots, 8 \) and residuals \( (e^{(k)}) \) for each outcome is Normally distributed and has zero mean.

For every explanatory variable (or its dummies), coefficients and standard errors were estimated. In order for each to be significant, the coefficient should be approximately twice its standard error, though exact levels of significance were ascertained using MLwiN’s “Intervals and Tests”. All explanatory variables were initially considered and those which were non-significant were removed.

Model selection is a complex process in its own right. Selection of sub-sets of covariates is not straightforward and can result in potential bias for both parameter estimates and their standard errors (Miller, 1990). However, with orthogonal variables (i.e. covariates with near-zero inter-correlations) and where the number of covariates is small, selecting only significant covariates does not generally yield problems of bias. Unless discussed in detail, it can therefore be assumed that all
models presented contain the sub-set of covariates that are significant (or borderline) without risk of bias. However, when the data set was reduced at T3, although the correlations between variables remained small and non-significant, covariate sub-set selection became less straightforward. Caution was therefore exercised in final model selection in these instances and this is discussed in more detail where this problem occurred.

**Dependent variable intercorrelations**

For each outcome variable the residual \(e^{(k)}\) provides a measure of the dispersion (i.e. variance), which assesses the differences between the predicted outcome (theoretical values from the model) and the actual outcome (observed values). For the null model, (i.e. where there are no explanatory variables), these variances are 1.0 (by design since all outcomes were centred and standardised). This variance reduces as explanatory variables are introduced to “explain” differences in each outcome. As there were multiple outcomes, the multivariate method also calculates the covariance for each pair of dependent variables. The covariance is directly related to the correlation coefficient. The MLwiN output provides variances (one for each outcome) and correlation coefficients (one for each pair of outcomes).

Correlations were evaluated for the null model (no explanatory variables) and compared with the final model (all significant explanatory variables) to provide additional information regarding the stability of the psychometric scales.

**Reversal of transformation and standardisation**

When the final model had been generated, the processes of standardisation and transformation were reversed in order to convert the findings back to the original dimensions of the outcome measures. However, it was pertinent to examine the results prior to this, as the transformed and standardised scales allowed the magnitude of the relative differences to be compared.
3.4.5 Miscellaneous modelling issues

A common methodological issue in longitudinal studies is that of how to handle missing data for the individuals who have failed to complete treatment in the time constraints of the study. The ideal method to account for this problem is to provide the subgroup mean for each of the missing respondents. However, this in itself becomes a problem with relatively small samples and several covariates. For example, the subgroup mean for black, female, class III may be derived from only one or two subjects. For this reason, the method cannot be recommended with small samples. In this study it was therefore necessary to omit follow-up subjects who had not completed treatment, with commensurate loss of statistical power. The implications of this are discussed in general with regard to appropriate study size in the final discussion section of this chapter.

In view of the fact that the number of respondents had decreased by T3, some of the explanatory variables were converted to binary forms rather than using all the categories. For example, ethnic group was categorised as Caucasian (reference group) or non-Caucasian, and severity of the problem was categorised as mild/moderate (reference group) or severe. However, combining subgroups that indicate differences in their influence on the outcome would be potentially erroneous. For instance, in study 2 when considering unit, unit 1 was the reference group and units 2 and 3 showed themselves to behave in a different manner with respect to unit 1 but in opposing directions with respect to each other. It was, therefore, felt that units 2 and 3 could not be combined and that there were insufficient subjects in the individual groups to allow them to remain in the model.

If insufficient numbers remained within individual subgroups to be used in the model in such circumstance, it was deemed more appropriate to exclude the covariate from the model altogether and to interpret the subsequent findings with the caveat that some of the residual variation in the model outcomes remained due to this covariate (but was unresolved due to small numbers). For the longitudinal study, the numbers of respondents from units 2 and 3 were small and considerably lower than those from
unit 1. For this reason, unit was excluded from subsequent longitudinal analyses and the model from study 2 was also re-run excluding unit (and using other multiple categorical variables in their binary form to make it comparable with the T3/T2 and T3/T1 analyses).

All models were estimated using “maximum likelihood” (ML) estimation procedures. For each model, the evaluated ML indicates how well the hypothesised model fits the data and a product of the ML procedure, which is used in model assessment, is $-2 \text{Log Likelihood} (-2 \text{LL})$. It becomes possible to compare the “fit” of a particular model with a perfectly fitting model with a likelihood of 1. Differences in $-2 \text{LL}$ between successive models (i.e. where covariates are added or removed) follows a $\chi^2$ distribution with the number of degrees of freedom (df) equal to the number of covariates removed or added. In order to establish whether a predictor (or set of predictors) should remain in a model involves two models, one with and one without the predictors. The $\chi^2$ value for the differences in $-2\text{LL}$ is thus used to determine if the models are significantly different (hence one is better).
3.5 STUDIES UNDERTAKEN

STUDY 1
A comparison of orthognathic patients and a control group of non-patients. The following explanatory variables were studied: group (experimental or control); age; gender and ethnic group.

STUDY 2
A study of changes between T2/T1 to establish whether there are any adverse effects due to pre-surgical orthodontics. The following explanatory variables were studied: age; gender; ethnic group; T2/T1 duration; antero-posterior (AP) skeletal relationship; the unit the patient was treated in; and patient-perceived severity of their dental/facial problems at presentation (T1). This necessarily involved only the experimental group.

STUDY 3
A longitudinal study of changes between T2/T1, T3/T2 and T3/T1. The following explanatory variables were studied: age; gender; ethnic group; time between questionnaires (T2/T1 duration and T3/T2 duration); antero-posterior (AP) skeletal relationship; and patient-perceived severity of their dental/facial problems at presentation. Again, this involved only the experimental group.
4.0 RESULTS

4.1 OVERVIEW OF ALL THREE STUDIES

Reliability testing
Table 63 presents the findings of the reliability testing (test-retest) for all outcome measures. The correlations were compared with those which have been previously quoted as acceptable. Not surprisingly, state anxiety gave a lower ICC than trait anxiety but this was anticipated due to the nature of the scale, in that it measures anxiety at that point in time and this may alter between tests. Social support (satisfaction) gave the poorest ICC and this will be mentioned further in the discussion. However, all values were considered to be sufficiently accurate to allow the outcome measures to be used in the study.

Demographic data
Tables 64 to 66 show the socio-demographic details for the respondents in each of the three studies. As anticipated, there was a greater proportion of females than males in the experimental group. This reflects the gender differences in patients requesting orthognathic treatment. The largest ethnic group was Caucasian, with fairly even numbers of other ethnic groups. The greatest number of respondents were drawn from unit 1 (Eastman Dental Hospital) which reflects the larger numbers of clinicians treating orthognathic patients in this unit compared with units 2 and 3, which are district general hospitals.

For Study 1, the mean age was 22.5 years (95% CI 21.7 to 23.3 years). The mean age for the 62 subjects who completed the T2 questionnaire was 21.9 years (95% CI 20.3 to 23.5 years) and 22.7 years (95% CI 20.6 to 24.8 years) for those who completed the T3 questionnaire.
The mean duration between the first and second questionnaires (T2/T1) was 17 months with 95% confidence intervals of 15 to 19 months (Table 65) and between the second and third questionnaires (T3/T2) was 8 months with 95% confidence intervals of 7 to 9 months (Table 66).

There were more patients with class III malocclusions than class II malocclusion (24 class II and 36 class III for T2 data; 15 class II and 26 class III for T3 data). One respondent had an anterior open bite on a class 1 skeletal base.

The majority of respondents perceived their dental/facial problems at the start of treatment as moderate or severe (23 and 30 respectively for T2 data) with only 9 stating that their problem was mild. These values were established from the VAS, with 0 to 3.3cm categorised as mild; 3.4 to 6.6 as moderate and 6.7 to 10cm as severe.
Table 63  
Reliability of the questionnaires.

<table>
<thead>
<tr>
<th></th>
<th>Pearson correlation</th>
<th>ICC</th>
<th>Suggested acceptable correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>0.80**</td>
<td>0.78</td>
<td>0.16-0.62 (Spielberger et al., 1983)</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.88**</td>
<td>0.87</td>
<td>0.65 - 0.86 (Spielberger et al., 1983)</td>
</tr>
<tr>
<td>Depression</td>
<td>0.93**</td>
<td>0.92</td>
<td>&gt;0.90 (Beck et al., 1961)</td>
</tr>
<tr>
<td>Social support (number)</td>
<td>0.80**</td>
<td>0.80</td>
<td>0.90 (Sarason et al., 1983)</td>
</tr>
<tr>
<td>Social support (satisfaction)</td>
<td>0.61**</td>
<td>0.60</td>
<td>0.83 (Sarason et al., 1983)</td>
</tr>
<tr>
<td>Self esteem</td>
<td>0.93**</td>
<td>0.93</td>
<td>0.85-0.92 (Rosenberg, 1965)</td>
</tr>
<tr>
<td>Body image</td>
<td>0.89**</td>
<td>0.89</td>
<td>Not quoted</td>
</tr>
<tr>
<td>Facial body image</td>
<td>0.81**</td>
<td>0.81</td>
<td>Not quoted</td>
</tr>
</tbody>
</table>

**p<0.01
Table 64  
T1: Comparison of experimental and control group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Gender</th>
<th>Age (years)</th>
<th>Ethnic group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>Male</td>
<td>71</td>
<td>Caucasian^R</td>
</tr>
<tr>
<td>Experimental</td>
<td>Female^R</td>
<td>110</td>
<td>22.5 (21.7 to 23.3)</td>
</tr>
<tr>
<td>Control^R</td>
<td>Male</td>
<td>96</td>
<td>Black</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>22</td>
<td>Other</td>
</tr>
</tbody>
</table>

^R indicates the reference category.

Table 65  
T2: Respondent details (n=62).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (years)</th>
<th>Ethnic group</th>
<th>T2/ T1 duration (months)</th>
<th>Unit</th>
<th>A-P skeletal base</th>
<th>Patient perceived severity at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female^R</td>
<td>Mean (95% CI)</td>
<td>Caucasian^R</td>
<td>Non-Caucasian</td>
<td>Mean (95% CI)</td>
<td>1^R</td>
</tr>
<tr>
<td>23</td>
<td>39</td>
<td>21.9 (20.3 to 23.5)</td>
<td>44</td>
<td>18 (Asian 6 Black 4 Other 8)</td>
<td>17 (15 to 19)</td>
<td>37</td>
</tr>
</tbody>
</table>

NB: One respondent had an anterior open bite only and a class I base.

^R indicates the reference category.
Table 66  T3: Respondent details (n=42).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (years)</th>
<th>Ethnic group</th>
<th>T2/T1 duration</th>
<th>T3/T2 duration</th>
<th>Unit</th>
<th>A-P skeletal base</th>
<th>Patient perceived severity at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female (^R)</td>
<td>Mean (95% CI)</td>
<td>Caucasian (^R)</td>
<td>Mean (95% CI) in months</td>
<td>1(^R)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>24</td>
<td>22.7 (20.6 to 24.8)</td>
<td>29</td>
<td>13</td>
<td>16 (14 to 18)</td>
<td>8 (7 to 9)</td>
</tr>
</tbody>
</table>

NB: One respondent had an anterior open bite only and a class I base.
\(^R\) indicates the reference category.
4.2 STUDY 1: A COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS

Tables 67 to 71 present the data for Study 1 which compared orthognathic patients with a control group of non-patients.

Table 67 presents the values for each of the eight outcome measures before transformation and standardisation were undertaken. The method of transformation used in each case is also indicated. Three of the eight outcome measured did not require transformation.

Table 68 shows the significant explanatory variables for each outcome measure. The final model, with regression coefficient estimates and standard errors for significant variables, is presented in Table 69. Of all the explanatory variables, group (patient or control) and gender were found to have the greatest influence on the outcome measures. Facial body image showed the greatest number of significant variables (group, gender, age and ethnic group) of all the measures. Table 70 presents the regression coefficients for the final model when converted back to the original outcome measure scale.

Correlations for the null and final models are presented in Table 71. It may be noted that there are only small variations between the two models for any of the correlations which lends support to the cultural and demographic stability of these outcome measures.
Table 67  Study 1  Comparison of experimental and control group: values prior to transformation and standardisation.

<table>
<thead>
<tr>
<th>Index (possible range)</th>
<th>Experimental group</th>
<th>Control group</th>
<th>Before transformation</th>
<th>Transformation</th>
<th>After transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Mean (SD)</td>
<td>Range</td>
<td>Mean (SD)</td>
<td>Skewness</td>
</tr>
<tr>
<td>State anxiety (20-80)</td>
<td>20-77</td>
<td>41.7 (10.6)</td>
<td>20-65</td>
<td>36.4 (9.9)</td>
<td>0.725</td>
</tr>
<tr>
<td>Trait anxiety (20-80)</td>
<td>25-78</td>
<td>43.8 (11.9)</td>
<td>24-66</td>
<td>40.9 (10.7)</td>
<td>0.689</td>
</tr>
<tr>
<td>Depression (0-60)</td>
<td>0-39</td>
<td>9.4 (9.1)</td>
<td>0-22</td>
<td>6.6 (5.3)</td>
<td>1.608</td>
</tr>
<tr>
<td>Social support: number (0-9)</td>
<td>0.3-9.0</td>
<td>4.2 (2.0)</td>
<td>0-9</td>
<td>3.6 (2.4)</td>
<td>0.546</td>
</tr>
<tr>
<td>Social support: satisfaction (0-6)</td>
<td>1-6</td>
<td>4.9 (1.0)</td>
<td>1-6</td>
<td>5.1 (1.0)</td>
<td>-1.743</td>
</tr>
<tr>
<td>Self esteem (10-40)</td>
<td>11-37</td>
<td>21.3 (5.8)</td>
<td>10-34</td>
<td>19.0 (5.3)</td>
<td>0.404</td>
</tr>
<tr>
<td>Body image (27-135)</td>
<td>34-123</td>
<td>82.1 (17.6)</td>
<td>27-108</td>
<td>70.3 (18.4)</td>
<td>-0.248</td>
</tr>
<tr>
<td>Facial body image (10-50)</td>
<td>16-47</td>
<td>33.3 (6.2)</td>
<td>10-37</td>
<td>23.5 (6.8)</td>
<td>-0.156</td>
</tr>
</tbody>
</table>
Table 68

Study 1 Comparison of experimental and control group: significant parameters for each outcome variable.

<table>
<thead>
<tr>
<th>Group</th>
<th>Significant variables</th>
<th>Ethnic Group</th>
<th>Age</th>
<th>ns</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support: number</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support: satisfaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = p<0.05; ** = p<0.01; *** = p<0.001
Table 69

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Intercept</th>
<th>Group</th>
<th>Gender</th>
<th>Age</th>
<th>Ethnic group</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>-0.17 (0.09)</td>
<td>0.32 (0.10)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.00 (0.08)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Depression</td>
<td>0.00 (0.08)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Social support: number</td>
<td>0.07 (0.12)</td>
<td>0.36 (0.15)</td>
<td>-0.33 (0.15)</td>
<td>-</td>
<td>-0.46 (0.23) Asian -0.51 (0.25) Black -0.54 (0.32) Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support: satisfaction</td>
<td>0.11 (0.09)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.49 (0.23) Asian -0.30 (0.24) Black -0.52 (0.31) Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td>0.06 (0.09)</td>
<td>0.19 (0.10)</td>
<td>-0.47 (0.10)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>-0.06 (0.10)</td>
<td>0.49 (0.12)</td>
<td>-0.45 (0.12)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>-0.44 (0.09)</td>
<td>1.08 (0.10)</td>
<td>-0.23 (0.11)</td>
<td>0.016 (0.006)</td>
<td>-0.11 (0.10) Asian -0.09 (0.11) Black 0.38 (0.14) Other</td>
</tr>
</tbody>
</table>

Baseline values are those for Caucasian, female, controls, aged 22 years.
Table 7.0  Study 1  Comparison of experimental and control group: values presented are regression coefficients converted back to the original outcome measure scales (confidence intervals in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Group</th>
<th>Gender</th>
<th>Age (per 10 years)</th>
<th>Ethnic 1 (South Asian)</th>
<th>Ethnic 2 (Black)</th>
<th>Ethnic 3 (Other)</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>35.81</td>
<td>3.24</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(34.18, 37.50)</td>
<td>(1.20, 5.38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>40.83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(39.24, 42.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>5.76</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(4.95, 6.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support: number</td>
<td>3.74</td>
<td>0.84</td>
<td>-0.69</td>
<td>-</td>
<td>-0.95</td>
<td>-1.05</td>
<td>-1.10</td>
</tr>
<tr>
<td></td>
<td>(3.24, 4.28)</td>
<td>(0.17, 1.57)</td>
<td>(-1.25, -0.08)</td>
<td></td>
<td>(-1.77, -0.01)</td>
<td>(-1.92, -0.04)</td>
<td>(-2.20, 0.20)</td>
</tr>
<tr>
<td>Social support: satisfaction</td>
<td>5.26</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.49</td>
<td>-0.28</td>
<td>-0.52</td>
</tr>
<tr>
<td></td>
<td>(5.11, 5.40)</td>
<td></td>
<td></td>
<td></td>
<td>(-1.25, -0.08)</td>
<td>(-0.96, 0.09)</td>
<td>(-1.61, 0.02)</td>
</tr>
<tr>
<td>Self esteem</td>
<td>20.43</td>
<td>1.05</td>
<td>-2.66</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(19.38, 21.48)</td>
<td>(-0.03, 2.13)</td>
<td>(-3.77, -1.56)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>74.68</td>
<td>9.31</td>
<td>-8.53</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(70.84, 78.53)</td>
<td>(4.83, 13.79)</td>
<td>(-13.09, -3.98)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>24.53</td>
<td>8.84</td>
<td>-1.89</td>
<td>1.30</td>
<td>-0.90</td>
<td>-0.73</td>
<td>3.10</td>
</tr>
<tr>
<td></td>
<td>(23.08, 25.98)</td>
<td>(7.16, 10.51)</td>
<td>(-3.58, -0.20)</td>
<td>(0.34, 2.27)</td>
<td>(-2.51, 0.71)</td>
<td>(-2.50, 1.04)</td>
<td>(0.85, 5.35)</td>
</tr>
</tbody>
</table>

Baseline values are those for Caucasian, female, controls, aged 22 years.
Table 71  Study 1  Comparison of experimental and control group: correlations for the null model (and for the final model in italics in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>State anxiety</th>
<th>Trait anxiety</th>
<th>Depression</th>
<th>Social support: number</th>
<th>Social support: satisfaction</th>
<th>Self esteem</th>
<th>Body image</th>
<th>Facial body image</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.73 (0.73)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>0.56 (0.56)</td>
<td>0.74 (0.74)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support: number</td>
<td>-0.11 (-0.15)</td>
<td>-0.12 (-0.16)</td>
<td>-0.18 (-0.22)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support: satisfaction</td>
<td>-0.29 (-0.26)</td>
<td>-0.31 (-0.29)</td>
<td>-0.26 (-0.25)</td>
<td>0.26 (0.25)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td>0.54 (0.52)</td>
<td>0.71 (0.70)</td>
<td>0.61 (0.62)</td>
<td>-0.17 (-0.28)</td>
<td>0.38 (0.40)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>0.39 (0.33)</td>
<td>0.49 (0.47)</td>
<td>0.46 (0.45)</td>
<td>0.03 (-0.09)</td>
<td>0.22 (0.22)</td>
<td>0.58 (0.52)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>0.47 (0.42)</td>
<td>0.44 (0.46)</td>
<td>0.41 (0.41)</td>
<td>0.00 (-0.17)</td>
<td>0.26 (0.24)</td>
<td>0.51 (0.49)</td>
<td>0.82 (0.83)</td>
<td>1.00</td>
</tr>
</tbody>
</table>
4.3 STUDY 2: DOES PRE-SURGICAL ORTHODONTICS RESULT IN ADVERSE PSYCHOLOGICAL EFFECTS?

Tables 72 to 76 present the data for Study 2. Table 72 shows mean scores for each of the outcome variables at T2 and T1 and also the differences in scores between the two time points. For five of the six scales there was a small increase in mean values between T2 and T1, only self esteem showed no change. The distribution of the differences was such that transformation was not required.

Table 73 shows the significant variables and Table 74 presents the final model. There were only four significant variables in total: one for state anxiety (age p<0.001); two for trait anxiety (gender p=0.04 and unit p=0.009) and one for body image (patient-perceived severity p<0.001). In addition the intercept for body image was significant (p=0.05). Interestingly, at this stage there were no significant findings for facial body image.

Table 75 presents the regression coefficients for the final model when converted back to the original outcome measure scale and model correlations are shown in Table 76. The consistency of correlations between final and null models further reinforces the concept that these measures are culturally and demographically stable.
Table 72  **Study 2 (T2/T1 comparison):** Values for the changes in scores T2/T1.

<table>
<thead>
<tr>
<th>Index (possible range)</th>
<th>T1</th>
<th>T2</th>
<th>Change T2 - T1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=62</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>Mean (SD)</td>
<td>Range</td>
</tr>
<tr>
<td>State anxiety (20 - 80)</td>
<td>22-77</td>
<td>41.97 (10.79)</td>
<td>27-78</td>
</tr>
<tr>
<td>Trait anxiety (20 - 80)</td>
<td>25-78</td>
<td>44.65 (12.27)</td>
<td>24-79</td>
</tr>
<tr>
<td>Depression (0 - 60)</td>
<td>0-38</td>
<td>9.76 (9.13)</td>
<td>0-44</td>
</tr>
<tr>
<td>Self esteem (10 - 40)</td>
<td>11-37</td>
<td>21.50 (5.65)</td>
<td>12-38</td>
</tr>
<tr>
<td>Body image (27 - 135)</td>
<td>34-117</td>
<td>83.21 (16.70)</td>
<td>51-123</td>
</tr>
<tr>
<td>Facial body image (10 - 50)</td>
<td>16-46</td>
<td>33.35 (6.12)</td>
<td>22-46</td>
</tr>
</tbody>
</table>

NB: T2 minus T1 column does not exactly equate T2/T1 change as a small number of data sets were not complete.

**Scoring system:**
- State and trait anxiety: Higher scores = greater anxiety
- Depression: Higher scores = greater depression
- Self esteem: Higher scores = lower self esteem
- Body image/ facial body image: Higher scores = lower body image
Table 73  Study 2 (T2/T1 comparison): Significant parameters for each outcome variable.

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Gender</th>
<th>Age</th>
<th>Ethnic group</th>
<th>Duration of pre-op orthodontics (T2/T1)</th>
<th>Malocclusion</th>
<th>Unit/hospital</th>
<th>Severity of problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>ns</td>
<td>ns</td>
<td>***</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>ns</td>
<td>*</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>**</td>
</tr>
<tr>
<td>Depression</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Self esteem</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Body image</td>
<td>*</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>***</td>
</tr>
<tr>
<td>Facial body image</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
</tbody>
</table>

* = p<0.05; ** = p<0.01; *** = p<0.001
Table 74  Study 2 (T2/T1 comparison): Significant coefficients (at 5% level) for each outcome variable (values presented are regression coefficient estimates with standard errors in parentheses and p-values in italics).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Gender</th>
<th>Age</th>
<th>Unit/ hospital</th>
<th>Severity of problem at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>0.22 (0.12)</td>
<td>-</td>
<td>0.06 (0.02)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.17 (0.15)</td>
<td>-0.36 (0.17)</td>
<td>-</td>
<td>-0.39 (0.25) unit 2</td>
<td>0.41 (0.19) unit 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>←------------------------p=0.009----------------→</td>
</tr>
<tr>
<td>Depression</td>
<td>0.14 (0.13)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Self esteem</td>
<td>-0.04 (0.13)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Body image</td>
<td>-0.30 (0.15)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.89 (0.27) sev1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>←------------------------p&lt;0.001----------------→</td>
</tr>
<tr>
<td>Facial body image</td>
<td>-0.14 (0.13)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Baseline values are those for a Caucasian, female, 22 year old, treated in unit 1, with a class III malocclusion and perceive their problem to be severe. No significant findings for ethnic group; duration of pre-surgical orthodontics or malocclusion.
Table 75  
Study 2 (T2/T1 comparison): Values presented are regression coefficients converted back to the original outcome measure scales (95% confidence intervals in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Gender</th>
<th>Age (per year)</th>
<th>Unit/ hospital</th>
<th>Severity of problem at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>2.45</td>
<td>-</td>
<td>0.67</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.22, 5.11)</td>
<td>(0.22, 1.11)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>1.52</td>
<td>-3.21</td>
<td>-</td>
<td>unit 2</td>
<td>unit 3</td>
</tr>
<tr>
<td></td>
<td>(-1.16, 4.19)</td>
<td>(-6.24, -0.18)</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-3.48 (7.94, 0.98)</td>
<td>3.66 (0.27, 7.05)</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>0.88</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-0.76, 2.53)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td>-0.16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-1.16, 0.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>-3.52</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>sev1</td>
</tr>
<tr>
<td></td>
<td>(-7.04, 0.00)</td>
<td></td>
<td></td>
<td></td>
<td>sev2</td>
</tr>
<tr>
<td>Facial body image</td>
<td>-0.60</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10.45 (4.11, 16.79)</td>
</tr>
<tr>
<td></td>
<td>(-1.72, 0.51)</td>
<td></td>
<td></td>
<td></td>
<td>6.10 (1.88, 10.33)</td>
</tr>
</tbody>
</table>

Baseline values are those for a Caucasian, female, 22 year old, treated in unit 1, with a class II malocclusion and perceive their problem to be severe. No significant findings for ethnic group; duration of pre-surgical orthodontics or malocclusion.
<table>
<thead>
<tr>
<th></th>
<th>State anxiety</th>
<th>Trait anxiety</th>
<th>Depression</th>
<th>Self esteem</th>
<th>Body image</th>
<th>Facial body image</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.37 (0.50)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>0.48 (0.54)</td>
<td>0.60 (0.54)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td>0.35 (0.38)</td>
<td>0.48 (0.50)</td>
<td>0.59 (0.59)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>0.05 (0.00)</td>
<td>0.26 (0.35)</td>
<td>0.27 (0.27)</td>
<td>0.30 (0.37)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>0.08 (0.08)</td>
<td>0.17 (0.21)</td>
<td>0.05 (0.02)</td>
<td>0.33 (0.31)</td>
<td>0.63 (0.65)</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 76: Study 2 (T2/T1 comparison): Correlations for the six outcomes in the null model (final model in italics in parentheses).
Relationship between patient-perceived severity and body image index

Tables 73 to 75 show a significant relationship between patient-perceived severity and the body image index. Figure 40, below, illustrates this relationship graphically, with the x-axis representing the three categories of patient-perceived severity and the y-axis representing the standardised body image scale. A marked increase in body image score (i.e. reduction in satisfaction with body image) can be seen for those individuals who initially rated their problem as mild, with less change for the moderate group and a change in the opposite direction for the group who classified themselves as having a severe problem.

Figure 40  Study 2 (T2/T1 comparison): The relationship between patient-perceived severity (x axis) and the body image index (y axis).

Figure 41 looks in more detail at the T1 scores for each of the 27 individual items in the body image scale, in an attempt to establish why it was that the body image scale was affected by patient-perceived severity and not the facial body image scale. The red bars represent items which also formed the facial body image scale and it is noticeable that six out of the ten facial body image items were already scoring towards the top of the scale at T1.
Figure 41: Distribution of scores for the Body Image Index at T1.

Score at T1

Item number - ranked by increasing magnitude of score at T1

Items 5-12, 14 and 24 are the items which constitute the Facial Body Image Index (red bars).
4.4 **STUDY 3: LONGITUDINAL STUDY OF CHANGES IN PSYCHOMETRIC VARIABLES**

**Overall comparisons**
Table 77 presents the data for the longitudinal analysis (T2/T1; T3/T2 and T3/T1) using the multilevel multivariate regression technique and also the more commonly used the Wilcoxon signed-rank test. As anticipated, slightly different levels of significance were found between the two types of analyses.

**T2/T1 model**
Table 78 represents the new T2/T1 model with unit omitted and with ethnic group and patient-perceived severity converted to binary variables. This allowed comparison with the T3/T2 and T3/T1 analyses more readily. It is noticeable that when unit was omitted from the model, gender also ceased to have a significant effect (see Table 74 for original model). Otherwise, the findings were identical.

**T3/T2 comparison**
Tables 79 to 81 present the data for the T3/T2 comparison. Table 79 shows the changes between T3/T2 and shows that all values decreased at T3 compared with T2. The final model in Table 80 shows only a small number of variables having any significant effect on the outcome measures. These were, noticeably, for state anxiety (gender p<0.001; T3/T2 duration, p=0.06) and trait anxiety (gender p=0.01; T3/T2 duration p=0.03). Age also had a significant effect on facial body image (p<0.01). Also of note are the significant intercepts for five of the six outcome measures. Table 81 shows the regression coefficients for this model when converted back to the original outcome measure scale.
Figures 42 to 44 present the longitudinal changes for state (Figures 42 and 43) and trait (Figure 44) anxiety. The influences of gender, age and T3/T2 duration are apparent.

**T3/T1 comparison**

Tables 82 to 84 present the data for the T3/T1 comparison. Table 82 shows a reduction in the mean scores of all measures at T3 as compared with T1. Again, there were a small number of explanatory variables in the final model (Table 83), but this time affecting solely body image (gender $p=0.02$; ethnic group $p=0.03$) and facial body image (gender $p=0.02$; age $p=0.002$; malocclusion $p=0.001$). As in the T3/T2 comparison, of note are the significant intercepts for five of the six outcome measures. Table 84 presents the regression coefficients for this model when converted back to the original outcome measure scale.
Table 77  Study 3  Longitudinal analysis: p-values as derived from the Wilcoxon signed-rank test and from MLM null model.

<table>
<thead>
<tr>
<th></th>
<th>Wilcoxon’s p-value</th>
<th>MLM p-value in null model</th>
<th>Significant parameters in covariate model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T2/T1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State anxiety</td>
<td>0.13</td>
<td>0.07</td>
<td>Age</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.42</td>
<td>0.32</td>
<td>-</td>
</tr>
<tr>
<td>Depression</td>
<td>0.19</td>
<td>0.17</td>
<td>-</td>
</tr>
<tr>
<td>Self esteem</td>
<td>0.86</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>Body image</td>
<td>0.93</td>
<td>0.82</td>
<td>Severity</td>
</tr>
<tr>
<td>Facial body image</td>
<td>0.22</td>
<td>0.21</td>
<td>-</td>
</tr>
<tr>
<td><strong>T3/T2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State anxiety</td>
<td>0.02</td>
<td>&lt;0.01</td>
<td>Gender; T3/T2 duration</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.05</td>
<td>0.07</td>
<td>Gender; T3/T2 duration</td>
</tr>
<tr>
<td>Depression</td>
<td>0.08</td>
<td>0.11</td>
<td>-</td>
</tr>
<tr>
<td>Self esteem</td>
<td>&lt;0.01</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>Body image</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td>Facial body image</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>Age</td>
</tr>
<tr>
<td><strong>T3/T1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State anxiety</td>
<td>0.03</td>
<td>0.04</td>
<td>-</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.02</td>
<td>0.05</td>
<td>-</td>
</tr>
<tr>
<td>Depression</td>
<td>0.18</td>
<td>0.17</td>
<td>-</td>
</tr>
<tr>
<td>Self esteem</td>
<td>0.02</td>
<td>0.01</td>
<td>-</td>
</tr>
<tr>
<td>Body image</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>Gender; Ethnic group</td>
</tr>
<tr>
<td>Facial body image</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>Gender; Age; Malocclusion</td>
</tr>
</tbody>
</table>
Table 78  Study 3  T2/T1 model re-run excluding unit and with dummy variables converted to binary variables (values presented are regression coefficient estimates with standard errors in parentheses and p-values at 5% level in italics).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Gender</th>
<th>Age</th>
<th>Ethnic group</th>
<th>T2/T1 duration</th>
<th>Malocclusion</th>
<th>Severity of problem at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>0.22 (0.12)</td>
<td>-</td>
<td>0.07 (0.02)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.11 (0.12)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Depression</td>
<td>0.13 (0.13)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Self esteem</td>
<td>-0.04 (0.13)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Body image</td>
<td>0.30 (0.15)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.59 (0.17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=0.05</td>
</tr>
<tr>
<td>Facial body image</td>
<td>0.14 (0.13)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Baseline values are those for a Caucasian, 22 year old, female, with a T2/T1 duration of 17 months, a class III malocclusion, and perceive their problem to be mild/moderate.
Table 79  **Study 3 (T3/T2 comparison): Values for the changes in scores between T3 and T2.**

<table>
<thead>
<tr>
<th>Index (possible range)</th>
<th>T3/T2 differences: range</th>
<th>Change T3/T2: mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety (20 - 80)</td>
<td>-35 to 15</td>
<td>-6.45 (13.58)</td>
</tr>
<tr>
<td>Trait anxiety (20 - 80)</td>
<td>-34 to 29</td>
<td>-3.76 (12.92)</td>
</tr>
<tr>
<td>Depression (0 - 60)</td>
<td>-31 to 18</td>
<td>-2.32 (8.59)</td>
</tr>
<tr>
<td>Self esteem (10 - 40)</td>
<td>-8 to 6</td>
<td>-2.13 (3.74)</td>
</tr>
<tr>
<td>Body image (27 - 135)</td>
<td>-38 to 31</td>
<td>-11.49 (14.52)</td>
</tr>
<tr>
<td>Facial body image (10 - 50)</td>
<td>-21 to 6</td>
<td>-9.23 (6.06)</td>
</tr>
</tbody>
</table>

**Scoring system:**
- State and trait anxiety: Higher scores = greater anxiety
- Depression: Higher scores = greater depression
- Self esteem: Higher scores = lower self esteem
- Body image/ facial body image: Higher scores = lower body image
Table 80  Study 3 (Final T3/T2 model): Significant coefficients (at 10% level) for each outcome variable (values presented are regression coefficient estimates with standard errors in parentheses and p-values in italics).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Gender</th>
<th>Age</th>
<th>Ethnic group</th>
<th>T2/T1 duration</th>
<th>T3/T2 duration</th>
<th>Malocclusion</th>
<th>Severity at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>-0.93(0.17)</td>
<td>0.98(0.25)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.12(0.06)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>-0.58(0.16)</td>
<td>0.51(0.20)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.11(0.05)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>p=0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>-0.20(0.16)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td>-0.51(0.16)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>-0.80(0.17)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>-1.48(0.16)</td>
<td>-</td>
<td>-0.05(0.02)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>p&lt;0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Baseline values are those for a Caucasian, 22 year old, female, with a T2/T1 duration of 17 months, a T3/T2 duration of 8 months, a class III malocclusion and perceive their problem to be mild/moderate.
Table 81  Study 3 (T3/T2 comparison): Values presented are regression coefficients converted back to the original outcome measure scales (95% confidence intervals in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Gender</th>
<th>Age (per year)</th>
<th>Ethnic group</th>
<th>T2/T1 duration</th>
<th>T3/T2 duration (per month)</th>
<th>Malocclusion</th>
<th>Severity of problem at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>-12.63</td>
<td>13.31</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.63</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-17.30, -7.96)</td>
<td>(6.44, 20.18)</td>
<td></td>
<td></td>
<td></td>
<td>(-3.28, 0.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>-7.49</td>
<td>6.59</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-1.42</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-11.68, -3.31)</td>
<td>(1.36, 11.82)</td>
<td></td>
<td></td>
<td></td>
<td>(-2.73, -0.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>-1.72</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-4.50, 1.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td>-1.91</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-3.12, -0.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>-11.62</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-16.63, -6.60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>-8.97</td>
<td>-</td>
<td>-0.30</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-10.94, -7.00)</td>
<td></td>
<td>(-0.55, -0.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Baseline values are those for a Caucasian, 22 year old, female, with a T2/T1 duration of 17 months, a T3/T2 duration of 8 months, a class III malocclusion, and perceive their problem to be mild/moderate.
Figure 42  
State anxiety: changes over the longitudinal periods T2/T1 and T3/T2 (females).
Figure 43  
State anxiety: changes over the longitudinal periods T2/T1 and T3/T2 (males).
Figure 44: Trait anxiety: changes over the longitudinal periods T2/T1 and T3/T2.
Table 82  Study 3 (T3/T1 comparison): Values for the changes in scores between T3 and T1.

<table>
<thead>
<tr>
<th>Index (possible range)</th>
<th>T3-T1 differences: range</th>
<th>Change T3/T1: mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety (20 - 80)</td>
<td>-34 to 35</td>
<td>-4.12 (12.88)</td>
</tr>
<tr>
<td>Trait anxiety (20 - 80)</td>
<td>-32 to 32</td>
<td>-3.69 (12.47)</td>
</tr>
<tr>
<td>Depression (0 - 60)</td>
<td>-27 to 19</td>
<td>-1.64 (7.88)</td>
</tr>
<tr>
<td>Self esteem (10 - 40)</td>
<td>-13 to 9</td>
<td>-1.69 (4.46)</td>
</tr>
<tr>
<td>Body image (27 - 135)</td>
<td>-44 to 24</td>
<td>-10.61 (13.39)</td>
</tr>
<tr>
<td>Facial body image (10 - 50)</td>
<td>-44 to 6</td>
<td>-8.83 (6.36)</td>
</tr>
</tbody>
</table>

Scoring system:
State and trait anxiety    Higher scores = greater anxiety
Depression                 Higher scores = greater depression
Self esteem                Higher scores = lower self esteem
Body image/ facial body image Higher scores = lower body image
Table 83  Study 3 (Final T3/T1 model): Significant coefficients (at 5% level) for each outcome variable (values presented are regression coefficient estimates with standard errors in parentheses and p-values in italics).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Gender</th>
<th>Age</th>
<th>Ethnic group</th>
<th>T2/T1 duration</th>
<th>T3/T2 duration</th>
<th>Malocclusion</th>
<th>Severity at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>-0.38(0.15)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p=0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>-0.38(0.15)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p=0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>-0.20(0.16)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td>-0.37(0.16)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p=0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>-0.92(0.19)</td>
<td>0.63(0.27)</td>
<td>-</td>
<td>-0.48(0.22)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>p=0.02</td>
<td></td>
<td>p=0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>-1.80(0.18)</td>
<td>0.51(0.22)</td>
<td>-0.04(0.01)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.58(0.18)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>p=0.02</td>
<td>p=0.002</td>
<td>p=0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Baseline values are those for a Caucasian, 22 year old, female, with a T2/T1 duration of 17 months, a T3/T2 duration of 8 months, a class III malocclusion, and perceive their problem to be mild/moderate.
Table 8.4 Study 3 (T3/T1 comparison): Values presented are regression coefficients converted back to the original outcome measure scales (95% confidence intervals in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Gender</th>
<th>Age (per year)</th>
<th>Ethnic group</th>
<th>T2/T1 duration</th>
<th>T3/T2 duration</th>
<th>Malocclusion</th>
<th>Severity of problem at T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>-4.89</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-8.79, -1.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>-4.74</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-8.51, -0.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>-1.58</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-4.12, 0.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self esteem</td>
<td>-1.65</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-3.09, -0.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>-12.32</td>
<td>8.44</td>
<td>-</td>
<td>-3.43</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(-17.46, -7.18)</td>
<td>(1.13, 15.74)</td>
<td>(-12.38, -0.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>-11.45</td>
<td>3.24</td>
<td>-0.25</td>
<td>-</td>
<td>-</td>
<td>3.69</td>
<td>-</td>
<td>(1.38, 6.00)</td>
</tr>
<tr>
<td></td>
<td>(-13.76, -9.14)</td>
<td>(0.42, 6.07)</td>
<td>(-0.38, -0.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Baseline values are those for a Caucasian, 22 year old, female, with a T2/T1 duration of 17 months, a T3/T2 duration of 8 months, a class III malocclusion, and perceive their problem to be mild/moderate.
5.0 DISCUSSION

5.1 RELIABILITY AND VALIDITY TESTING

On the whole, the results of the test-retest study gave correlation coefficients which were within the range quoted for each of the individual instruments (Table 63). For example, Spielberger et al. (1983) suggested that coefficients between 0.65 and 0.86 should be achieved with their trait anxiety scale and Beck et al. (1961) recommended values over 0.90. Only social support (satisfaction) gave doubtful levels of reliability and this is discussed later.

Validity is an important issue when using any questionnaire with a population which is different from the one upon which it was developed. The correlations presented in Tables 71 and 76 show evidence that all of the outcome measures were culturally and demographically stable and provides validity for their use in this study.

5.2 USE OF MULTILEVEL MULTIVARIATE ANALYSIS

Multilevel multivariate regression was undertaken for several reasons:

- To optimise the modelling for each psychometric scale through consideration of all eight outcomes simultaneously.
- To account for residual demographic variation between the experimental and control groups in the study comparing experimental and control groups (since respondents had not been matched one for one) and to explore how differences were attenuated by demographic factors in all subsequent studies.

Results of multivariate analysis may reveal findings that differ from those of independent regression models for each outcome measure. In comparison to the independent models, multivariate analysis estimated the impact of the explanatory variables with greater confidence i.e. smaller standard errors. For example, in Study 1, the coefficient for group in the separate social support (number) model was not
significant \( p=0.06 \) whereas it was in the multivariate model \( p=0.01 \). It may be that there were insufficient subjects to achieve significance in the separate analysis and multivariate analysis improved statistical power. Therefore, multivariate analysis is advantageous in reducing Type II errors, where explanatory variables are wrongly assessed as non-significant due to insufficient data. In contrast, for separate regression analyses, the coefficient for group was significant for self esteem \( p=0.006 \) but was only of borderline significance in the multivariate technique \( p=0.06 \). Therefore, multivariate techniques also have strengths in avoiding Type I errors, where explanatory variables are incorrectly assessed as being significant due to chance or data artefact. Multivariate techniques are therefore statistically more efficient, with each regression model “borrowing strength” from the parallel process of multiple simultaneous regression models.

There are also a number of other benefits in the use of multivariate analysis. It is particularly suitable for exploring data where there are missing or incomplete responses (missing data in this context is different to that already discussed with respect to loss of subjects in the longitudinal analysis). This may be either by design, in an effort to improve response rates, or by accident, provided the data is missing at random. The multivariate technique also determines correlations between pairs of outcomes, even where some data are missing.

5.3 STUDY 1: COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS

The significant parameters for each outcome measure are shown in Table 68. Group was significant for state anxiety, social support (number), body image and facial body image, with borderline significance for self esteem (Table 69). Group (experimental or control) was retained in the model for self esteem although it was of borderline significance as this variable was of particular interest as a focus of the investigation.
In addition, the model did not exhibit any changes if group was removed for self esteem.

**Group**
Table 70 shows the results of the multivariate analysis with values converted back to the original outcome measure scales. The experimental group was likely to show a higher level of state anxiety than the control group, this value being on average 3.24 points higher when converted back to the original outcome scale. In addition, the experimental group was likely to report higher numbers of individuals in their social support network (mean difference of 0.84 people) and less satisfaction with their overall body image and facial body image (represented by body image/facial body image scores which were 9.31 and 8.84 points higher respectively). Self esteem was of borderline significance, with the experimental group showing lower self esteem (i.e. higher score). This may have been a "true" finding or may have related to underlying distributional differences (i.e. self selection).

**Demographics**
The age of the respondent was significantly associated with only one of the outcomes: facial body image. For every 10 years increase in age, dissatisfaction with facial body image increased by 1.3 points (Table 70). However, this is probably not of any clinical relevance since the magnitude of this covariate was small.

Gender was significantly associated with four outcomes: number of individuals in the social support network; self esteem; body image and facial body image. Males reported fewer individuals for social support (mean difference of 0.69 people) than females, who were also likely to report lower self esteem (indicated by higher psychometric scores) and less satisfaction with their body image and facial body image (also indicated by higher psychometric scores).
There were some significant differences between Caucasian and minority ethnic group respondents, even though not all individual minority ethnic group terms were significant. Ethnic group was significant for both the number of people in the social support network and satisfaction with this support. For social support (number), South Asian and black respondents showed significantly fewer people in their support network when compared with Caucasians. The coefficient for the remaining ethnic group ("others") did not reach significance. For social support (satisfaction), all minority ethnic groups showed less satisfaction with their social support but this attained significance only for the South Asian group (mean difference of 0.49; Table 70).

South Asian and black respondents reported greater satisfaction with facial body image than Caucasian subjects, but only the racial group labelled "other" attained any significant differences, having less satisfaction with their facial body image when compared with Caucasian subjects.

**General discussion**

The results showed that there were several parameters which affected the eight outcome variables. The only significant parameter affecting the difference in the state anxiety scores was whether the respondent was in the experimental or control group. As state anxiety is defined as transitory feelings of fear/worry, it is perhaps not surprising that the experimental group achieved a higher average score in this variable. In contrast, the dentofacial group did not have a significant effect with respect to trait anxiety. Thus, the experimental and control groups did not differ fundamentally in their stable tendency to respond to anxious situations, although they exhibited different transitory levels of anxiety. This supports the concept that the two forms of anxiety should be assessed separately.

The dentofacial group was not significant for the Beck Depression Inventory. It is encouraging that orthognathic patients do not show higher levels of depression before
treatment. Therefore, the depression experienced by some patients following surgery is likely to be reactive depression as a result of the treatment rather than endogenous depression. However, when the data are looked at in isolation, the mean score for the BDI at T1 was 9.4 for the experimental group and 6.6 for the control group. The usual scoring guides based on normative data suggest that scores of 0-9 suggest no depression and 10-15 indicate mild depression (Bowling, 1995), in which case the experimental group showed very mild levels of depression.

The report of higher numbers of individuals in the support network of the experimental group may be relevant during treatment, even though satisfaction with social support did not show significant differences. Holman et al. (1995) noted the importance of social support particularly in the post-operative period and this may influence satisfaction following treatment.

Although group showed only borderline significance for self esteem values, this is an interesting area which could be studied further, possibly with a larger sample size to reduce the possibility that this observation was borderline only due to lack of statistical power due to sample size. The parameters which affect self esteem are still largely unknown.

The dentofacial group findings regarding body image are not surprising. It is reasonable to assume that a patient who is prepared to go through extensive treatment to change their appearance may have lower satisfaction with their body image (i.e. higher body image scores), particularly with respect to their facial features. The facial body image variable revealed the largest number of significant independent variables in the regression model. Satisfaction with facial body image decreased with age. It may therefore be interesting to study a group of older patients requesting orthognathic surgery and determine whether this is generally the case. This may also have some bearing on the success of outcome when operating on older patients.
The results regarding gender showed that males reported fewer individuals in their social support network, but interestingly satisfaction with support was not significantly different. Females showed lower self esteem and reduced satisfaction with body image. Kiyak et al. (1981) reported that aesthetic improvement was important for both males and females, but this study suggests that females are starting from a lower baseline value than males.

Ethnic group was significant with respect to the number of individuals in the social support network, with minority ethnic groups reporting a lower number of individuals with respect to the Caucasian group. Satisfaction with social support was also lower in the minority ethnic groups, with the South Asian group experiencing this the most. This is important because it may be that certain ethnic group patients require additional support during treatment and the clinician should be particularly vigilant. It is also of interest that “other” racial groups reported significantly lower satisfaction with facial body image than Caucasian respondents.

**Model correlations**

The most notable correlations (Table 71) were for trait and state anxiety (0.73); depression and trait anxiety (0.74); self esteem and trait anxiety (0.71); self esteem and depression (0.61) and body image and facial body image (0.82). These did not alter significantly after modelling. This, in conjunction with the knowledge that matching was generally good, suggests that the chosen psychometric variables are demographically and culturally stable. In other words, the measures used were robust across the age, gender and ethnic group composition of the target study groups. This is encouraging as one of the criticisms sometimes levelled at psychometric tests is that they are not culturally diverse. Social support (number) showed noticeable variation between the null and covariate models, suggesting this scale is less stable. The social support questionnaire was also commented on by a number of respondents as being onerous to complete. These two factors, combined with the low ICC for reliability of
reliability of the social support satisfaction component, led to the decision not to use this scale in Studies 1 and 2.

There were a number of negative correlations relating to social support. As social support (number and satisfaction) increased, state and trait anxiety and depression reduced. This is an encouraging finding for surgical patients in that, if their social support network can be encouraged to play a part in their treatment, they may show less anxiety and depression both during and after treatment. As social support (numbers) increased, self esteem, body image and facial body image values reduced. However, it must be borne in mind that all the negative correlations were very low and would account for only a small percentage of the variation seen.

Even though state and trait anxiety were measuring different aspects of anxiety, the high correlation between state and trait anxiety is perhaps not surprising. Neither is the correlation between body image and facial body image, since the latter is a sub-component of the body image scale. However, the correlations between self esteem and trait anxiety (0.71) and self esteem and depression (0.62) were interesting. As trait anxiety or depression increased, self esteem was found to decrease and vice versa. Very little is known regarding inter-relationships between these outcome measures and it would be interesting to look into this further.

Reversal of data standardisation
An important feature of standardising the data is seen when comparing the relative impact of the explanatory variable. For instance, the coefficients for group effects within both body image scales were approximately equal (Table 70). However, for the same results in the standardised dimension (Table 69), the group coefficient was twice as large for facial body image as for overall body image, and differences in comparison to other scales were also changed. This is because the underlying distribution of both scales differed (Table 67), the body image exhibiting a wider range of values than facial body image. What initially seemed to be similar in absolute
numerical differences between patients and controls was in fact a much larger effect for facial body image than overall body image. Studying the findings in terms of both standardised and non-standardised results is therefore essential in interpreting appropriately the role of the explanatory variables.

**Use of multivariate (multilevel) regression**

The reason multivariate regression was undertaken was two-fold: to optimise the modelling for each psychometric scale through consideration of all eight models simultaneously and to account for any residual demographic differences between the experimental and control groups since respondents had not been matched one for one. Therefore, independent demographic variables which were found to be significant may have resulted from poor matching. Alternatively, differences might be inherent demographic contrasts measured by the associated psychometric scale. For instance, age was shown to be significant only for facial body image though this was probably not clinically significant. In addition, gender differences reflected anticipated differences between males and females. Combined, these findings suggest good matching for age and gender. However, it was anticipated that matching would be more difficult for ethnic groups, where there is little known regarding ethnic group differences among the psychometric measures investigated. It is likely that there was a combination of both poorer ethnic group matching and some cultural variability among the psychometric scales.

This form of analysis is likely to have many applications in the future for analysis of dental data and previous studies might have benefited from this approach. It could be utilised, for example, in studies like that undertaken by Kiyak et al. (1986) which investigated differences in psychometric outcomes for orthognathic patients, those patients who decided against treatment and those who decided to undergo orthodontics only. All outcomes could be looked at simultaneously and demographic differences also accounted for.
5.4 STUDY 2: DOES PRE-SURGICAL ORTHODONTICS RESULT IN ADVERSE PSYCHOLOGICAL EFFECTS?

Many studies have looked at changes in psychological profile following orthognathic treatment but the baseline which has been used for pre/post treatment comparisons is typically in the immediate pre-surgery phase when pre-surgical orthodontic treatment has been completed (Kiyak et al., 1982, 1986; Flanary et al., 1990; Barbosa et al., 1993; Finlay et al., 1995). There is very little information regarding any potential psychological effects of pre-surgical orthodontics. The aim of this section of the study was therefore to determine whether any psychological changes occurred during the pre-surgical phase as a result of orthodontic treatment and whether it was acceptable to use the pre-surgical stage as a baseline.

Model intercepts were of particular importance because they represented the mean change in each outcome due to the effect of the pre-surgical orthodontics, having accounted for all other explanatory variables. Results are presented in Tables 72 to 76.

State Anxiety
Age was significantly associated with changes in state anxiety (Table 75). For every year increase in age, state anxiety levels increased by 0.67 points when converted back to the original scale (p<0.001). This could have implications when treating older patients as they may become more anxious prior to surgery and require more support from the clinicians involved in their care. Although the intercept was not significant in the null model, what moderate effect there was due to treatment was attenuated by age.

Trait Anxiety
Trait anxiety is the in-built response to anxious situations and changes were not expected in the same way as for state anxiety even with a time interval of between 10
and 29 months between the two questionnaires. However, changes in trait anxiety were noted. Although the developers of these scales claimed that state and trait anxiety are measured separately, it is now widely believed that this separation between the two forms of anxiety is not clear cut. Indeed, a number of other studies have also found changes in trait anxiety following treatment (Feinmann, Personal communication).

Changes in trait anxiety varied significantly by gender and unit (Tables 74). Although females showed a small non-significant increase in trait anxiety (0.17, ns), males actually showed a decrease in state anxiety (-0.19, p=0.04). This equated to 1.52 and -1.69 points, respectively, when converted back to the original scale (Table 75). However, the level of significance for gender was only p=0.04 and later analyses (Table 78) showed that when unit was removed from the model, gender was no longer significant. Hence, this may have been a Type I error and it appears that the effect of gender must be treated with some caution. In addition, the unit in which the patient was treated affected trait anxiety. Although this was significant overall, it was changes amongst those subjects from unit 3 which attained significance with respect to the base unit (unit 1). Changes at unit 2 did not significantly differ from changes at unit 1 and no significant changes due to orthodontics per se were observed. Furthermore, it should be noted that when initial trait anxiety was looked at for each of the units independently, unit 3 showed a tendency for higher pre-treatment values than those individuals in unit 1 (p=0.11) and 2 (p=0.22). Although this was not significant, the sample size for unit 3 was small (16 in total), so this may have been a sample size effect. Consequently, there may have been some underlying differences in the patient group selected at unit 3, although this did not affect the conclusions regarding the role of pre-surgical orthodontics.

**Body Image**

The respondents' perceived severity of their dental/facial problems (at T1) showed significant changes for body image scores, but interestingly not for facial body image.
When the initial and final models were compared it was noted that the significance of this explanatory variable was elevated in the final model, when all other non-significant explanatory variables had been removed. Therefore, the effect may be lower than the value $p<0.001$ would suggest. The intercept (which represents the mean changes in outcome due to pre-surgical orthodontics, having accounted for explanatory variables) was weakly significant for body image ($p=0.05$) but was not significant in the null model.

Table 74 and Figure 40 show that those individuals who perceived their problem to be mild initially had, on average, a greater increase in body image score (i.e. a greater reduction in their body image satisfaction) as a result of intervention. Those who reported their problem to be moderate also experienced a reduction in body image satisfaction although to a lesser extent. Interestingly, those who reported a severe problem at T1, showed a slightly reduced body image score (i.e. a small improvement in their satisfaction with body image). This may have been in anticipation of the changes expected due to surgery in the near future. Therefore, those individuals who rate their problem as mild or moderate initially should be given additional warnings that there may be worsening of how they perceive their problems as treatment proceeds. It is, however, encouraging to note that the reduced satisfaction with body image did not appear to affect other outcomes such as anxiety or depression.

It is interesting that body image rather than facial body image was affected most. Therefore, individual item responses at T1 were also looked at for the complete body image scale (Figure 41). It was noted that those items in the facial body image index, when studied individually, showed a significantly different mean change as compared with the items not in the facial body image component. In addition, Figure 42 illustrates that the facial body image items tended to score at the upper end of the scale in the T1 questionnaire and this may create a "ceiling" effect, where facial body image items have a reduced opportunity to increase. Respondents may then express their dissatisfaction as a general dissatisfaction affecting other body parts as well as the facial area.
CHAPTER 4  DISCUSSION

**General discussion**

Reassuringly, duration of pre-surgical orthodontics did not have a significant influence on any of the outcome measures. The range of treatment times varied between 10 and 29 months and it was thought that those with longer treatment times may show a greater increase in scores for anxiety and depression, but this was not substantiated. This, in itself, is a very important finding.

As in the comparison of the experimental and control group, correlations in the null model differed very little from those in the final model (Table 76), lending further support to the conclusion that the chosen psychometric variables are demographically and culturally stable. Interestingly, correlations for state and trait anxiety showed the greatest variation between the null and covariate models. This may help to explain some of the findings with respect to these scales in the longitudinal analyses.

When the multivariate/multilevel results were compared with the separate single level regression models for each outcome variable, some differences were noted. For state anxiety, age showed p-values of 0.005 in single level and <0.001 in multilevel analyses. For trait anxiety, gender and unit were both significant as in the multivariate analysis but both appeared to exhibit higher levels of significance in the single level analysis (i.e. gender: p=0.01 in single level and 0.04 in multilevel analyses). The same applied for body image, in which perceived severity was not significant in the separate regression model but was in the multivariate analysis. This was probably an example of a *Type II* statistical error.

The most important finding in this comparison of changes between T1 and T2 was that intervention, in the form of orthodontic treatment, had minimal effect on the chosen psychometric outcome variables. Previous researchers were probably justified in using the pre-surgical measurement as a baseline when investigating psychometric outcomes. The mean changes as a result of orthodontic intervention, when all other explanatory variables had been accounted for, were significant only for the body image index and this was only when patient-perceived severity was also considered.
Those who reported the least severe conditions at T1 showed the greatest increase in body image score (i.e. reduced satisfaction with body image) between T1 and T2. One additional finding was that older patients requesting orthognathic treatment may need additional support prior to surgery due to the tendency for state anxiety levels to increase more in older patients.

5.5 STUDY 3: LONGITUDINAL DATA ANALYSES

In the models for each of the time intervals in question, the intercept represented the mean change in outcome due to the intervention for all baseline subjects. For T3/T2 and T3/T1 comparisons, a number of the intercepts were found to be significant. For the T3/T2 comparison, four of the outcome measure intercepts were significant in the null model with an additional outcome measure achieving significance in the final model. For the T3/T1 model, intercepts for five outcome measures achieved significance in both models.

5.5.1 Comparison of data between the start of treatment and pre-surgery (T2/T1)

Table 78 shows the model for T2/T1 when it was re-run with unit excluded and with gender and severity as binary variables, thus allowing comparison with the T3/T2 and T3/T1 models more readily. It can be seen that gender no longer remained in the model for trait anxiety when unit was removed. This suggests that gender was of questionable “true” significance in its own right but was most likely confounded by unit, exhibiting a group selection effect at different units. It therefore appears that the role of gender in the earlier model may have been artefactual. Otherwise, all other findings remained consistent.
Although unit was not included in subsequent analyses for the reasons described earlier, it is accepted that there may be some residual differences due to unit.

5.5.2 Comparison of data between pre-surgery and the end of treatment (T3/T2)

For the T3/T2 final model, the intercept was significant for five of the six outcomes: state anxiety; trait anxiety; self esteem; body image and facial body image. This represented those variables which showed positive changes as a result of treatment, when all other explanatory variables had been taken into account (Table 80).

State anxiety
Changes in state anxiety varied significantly with gender, with females showing a reduction in anxiety (-0.93, p<0.001) and males showing a very small increase in score (0.05, p<0.001). It may be that males found the post-operative period more traumatic and stressful than females and therefore failed to show the same reduction in anxiety levels which females showed. Alternatively, males may show less fluctuation in their response to stressful situations.

The duration of time between T2 and T3 was of borderline significance (p=0.06). As the T2/T3 duration increased, anxiety levels reduced. This is discussed more fully in the following paragraph. Interestingly, in the full model, the p-value for T3/T2 duration was higher (p=0.04) than in the final model. This reflects that potential bias is occurring when selecting the final sub-set of covariates and it thus becomes less clear what the “true” size or significance of the modelled effect are. This problem developed due to reduced numbers in this model.
Trait Anxiety

Changes in trait anxiety also varied significantly with gender. Again, females showed a reduction in anxiety (-0.58, p<0.001), with males showing a slight decrease of much smaller magnitude (-0.07, p=0.01). The same issues discussed previously regarding whether this scale truly measures a trait are again pertinent.

The duration of time between T2 and T3 significantly influenced trait anxiety (p=0.03), with greater reduction in anxiety levels noted in those who had been in treatment for longer. This may reflect the fact that patients became resigned to the length of time that treatment was taking and their anxiety therefore reduced. It is also possible that anxiety levels peaked at an earlier stage and then reduced. In order to investigate this further, higher order terms for T3/T2 duration were introduced into the model: squared; cubed and quartic terms. This confirmed that there was a non-linear relationship between T3/T2 duration and trait anxiety and this influence will be investigated more fully in future investigations. Interestingly, the same did not apply to state anxiety, but it should be noted that the p-value in the final model was borderline.

Figures 42 to 44 show the longitudinal changes between T2/T1 and T3/T2 for both state and trait anxiety. Figure 42 and 43 show the changes for state anxiety. The graphs for females and males are presented separately for clarity. In this instance, there was an age effect only between T2/T1 and a gender and T3/T2 duration effect between T3 and T2. There are several points to note from these figures. The first is the greater treatment effect experienced by females with respect to males. Males show only small changes in state anxiety during treatment, with a slight increase following treatment. In contrast, females showed an increase in the pre-surgical period and a marked decrease following treatment. The second issue is the obvious age effect for both gender groups, in particular, older patients tended to have elevated anxiety levels in the T2/T1 period. The effect of the T3/T2 duration is also apparent. These findings have important clinical implications, particularly with respect to age as older patients may need special monitoring, and in some cases counselling, during
their treatment. Figure 44 shows the changes for trait anxiety (based on the original T2/T1 model: Table 75). The gender effect between T3/T2 was further influenced by T3/T2 duration, with those who were in treatment longer showing greater reductions in score (i.e. less anxiety).

For all figures it is important to note that 95% confidence intervals are not plotted, therefore the precision of the values should be interpreted with caution. The figures do, however, give an indication of the pattern of changes occurring.

**Self esteem**
Interestingly, although there were no significant parameters associated with self esteem, the intercept was significant (p<0.01). The fact that there were no parameters retained in the covariate model, illustrates the complex nature of self esteem. There may be interrelationships between self esteem and some of the other outcome measures which could go some way to explaining this finding. For example, the significant improvement in self esteem may be related to the improved body image and reduced anxiety. The question as to whether self esteem is improved by interventions such as orthognathic surgery has long been the subject of debate. This finding is therefore encouraging and may form the basis for future work in this area.

**Body Image**
The final model had no explanatory variables associated with changes in body image. It must be noted, however, that the intercept was significant at p<0.001.

**Facial Body Image**
Changes in facial body image varied significantly with age. Older patients showed a greater reduction in facial body image score than younger patients (i.e. their satisfaction with their facial body image increased more than for younger
respondents). It may be that older patients have “suffered” for longer as a result of their dentofacial deformity and therefore immediately perceive the changes more positively. The small changes which were noted, however, may not be of clinical significance (-0.30 points for every years increase in age when converted back to the original scale). Nevertheless, the intercept was highly significant (p<0.001) indicating significant improvement in facial body image.

**Interaction of gender and malocclusion**

In the full model, both gender and malocclusion exhibited a borderline level of significance which precluded them from being considered by the final model. However, it was thought feasible that there may be some interaction between gender and malocclusion which influenced the outcome. It is possible that female patients with a class III malocclusion show the greatest improvement in their body image because it is seen as an unattractive, masculine trait. For this reason, a new variable was introduced into the T3/T2 model which was the product of the gender and malocclusion variables. This variable accounted for the possible effect of the interaction of gender and malocclusion combined, in addition to their separate independent effects. However, the interaction variable was not significant for any of the outcomes and the model was not improved. Therefore the hypothesis was rejected.
Figure 45  Interaction variable representation.

<table>
<thead>
<tr>
<th>Female (reference)</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>III (reference)</td>
<td>Baseline case</td>
</tr>
<tr>
<td>II</td>
<td>Role of malocclusion independent of gender</td>
</tr>
</tbody>
</table>

5.5.3  Comparison of data between the start and the end of treatment (T3/T1)

The final model for T3/T1 is shown in Table 83. As for the T3/T2 model, the intercepts were significant for five of the six outcomes: state anxiety; trait anxiety; self esteem; body image and facial body image. There were no significant explanatory variables for state or trait anxiety, depression or self esteem.

Body Image
Changes in body image score varied significantly with gender and ethnic group. Females showed a greater reduction in body image score (i.e. greater satisfaction with body image) than males although both showed a reduced score. In addition, non-Caucasian respondents showed a greater reduction in score (i.e. a greater satisfaction with body image) than Caucasians. The intercept was, as before, significant (p<0.001).
Facial Body Image
Changes in facial body image score varied significantly with age, gender and malocclusion and the intercept was significant (p<0.001). Older patients showed a greater reduction in facial body image score than younger patients (i.e. an increased satisfaction with facial body image). However, these age changes were small and may not be of clinical significance. Class III respondents showed a greater reduction in facial body image score (i.e. a greater satisfaction with body image) than class II individuals, although both showed marked improvements. These changes may have important implications and they introduce the interesting question as to whether class III malocclusions are considered to be more disfiguring than class II malocclusions. This may be an interesting area for future research.

Gender also had significant effects, with males showing less reduction in their facial body image score than females. However, this result should be treated with some caution as the effect was much lower in the full model where its effect was being diluted in the presence of other non-significant variables. It is therefore unclear as to the magnitude of the gender effect, due to potential bias within this model which arises from small sample size (n=42).

Interaction of gender and malocclusion
The interaction term was introduced in the same way described for T3/T2 but, again, was not significant for facial body image.
5.6 OVERALL SUMMARY OF FINDINGS

5.6.1 Comparison of findings with other studies

It is impossible to compare this study directly with previous studies in view of the
different type of study design and analysis. Comparisons are further complicated by
the fact that different instruments are used in different studies. In addition there are
few studies which have looked at psychometric outcomes on more than two
occasions (usually before and after surgery). The longitudinal studies by Kiyak and
colleagues (Kiyak et al., 1981, 1982a, 1982b 1984, 1985, 1986) and Flanary et al.
(1990) are amongst the few studies to monitor orthognathic patients longitudinally,
although with T1 questionnaires being completed prior to surgery rather than at the
very start of treatment. The present study showed that, overall, there were a number
of positive changes as a result of orthognathic treatment. This finding compares with
other studies which have noted the same improvements (Kiyak et al., 1984; Flanary et
al., 1990).

This study found increases in body image and facial body image satisfaction at T3 as
compared with T1 and T2 (mean time post surgery 8 months). In contrast, Kiyak et
al. (1984) noted that satisfaction with body image decreased at the 9 months post-
operative stage and then increased again (beyond pre-treatment levels) by 24 months.
However they did draw attention to the fact that a number of patients were still
wearing fixed appliances at this stage. This finding lends support to the suggestion
that post-operative orthodontics should be kept to a minimum. In the same paper,
Kiyak et al. noted that self esteem remained significantly lower post-operatively than
at the start of treatment but hypothesised that levels were raised at T1 in anticipation
of surgery. A similar hypothesis was suggested by Cunningham et al. (1996) in their
paper. Again, in contrast, this study noted positive effects on self esteem with
statistically significant increases following treatment. This is an interesting area which
may form the basis for further study. It is still a matter for debate as to whether
orthognathic treatment per se is the major determinant of changes in self esteem
(Kiyak et al., 1986).
The positive changes in anxiety noted in this study were also noted by Kiyak et al. (1985) who found increased tension levels over the time period immediately prior to surgery and this was followed by a decrease to pre-surgical levels when fixation was removed. Obviously, the use of fixation may have further influenced these findings. Reassuringly, there were no significant findings with respect to depression in this study: patients did not show significantly different levels of depression when compared with members of the control group nor were there significant longitudinal changes. Kiyak et al. (1985) noted increases in depression in the immediate postsurgery phase (4-6 weeks after) and it may be that any reactive depression experienced by the patients in this study had passed by the time they completed their third questionnaire.

5.6.2 Statistical methods

This study has looked at changes in psychometric outcomes using a very different approach to that which has been used previously. This has allowed an insight into factors which may influence changes in these outcome measures during treatment. Rather than looking at simply the mean change in scores between two groups or two time intervals, explanatory variables were also introduced.

This methodology used a multilevel modelling framework allowing the study data to be explored optimally. Much of the data collected in medicine and dentistry is of a multilevel nature and by failing to analyse it appropriately, important findings may be overlooked (Type II errors) or findings may be attributed with significance when this is not the case (Type I errors).

One of the main problems encountered in this study was reduction in sample size. Despite this, sample size was comparable with other studies of this type. For example, Kiyak et al. (1982) reported findings for 55 patients but the respondents were studied over only a 10 month period and, in a later paper, the same group reported complete...
data for 46 of the 74 patients recruited for their longitudinal study (Kiyak et al., 1984).

MLM has clearly increased statistical power for this study, but the problem of loss of subjects to follow-up during the study period has become apparent and the multivariate regression technique, whilst able to improve matters, could not compensate totally for loss of follow-up at T3. This is an important observation since studies below a certain threshold of size (around 40 to 45) are likely to suffer similar problems to those observed in this study where models at T3 were sometimes less robust and model selection became a matter of careful judgement.

The implications of this are that studies which do not have considerably more subjects and/or do not adopt the multilevel framework, are likely to be less reliable than they could be. This emphasises the importance of both good study design (i.e. adequate sample size) and good analytical methods.

One of the most important aspects of this type of research is the interface between clinicians and statisticians. Only by collaborating in this way can clinicians learn to use new techniques which are necessary for appropriate data analysis.

5.6.3 Clinical implications

The first part of the study confirmed that there are some differences between the orthognathic patient and members of a control group. Some of these differences may have implications during treatment (for example, the patient may have pre-existing higher levels of anxiety). Comparison of the groups was allowed by virtue of the multivariate multilevel technique which accounted for demographic differences between the two groups where one-to-one matching was not feasible. This is an important use of the technique in medicine and dentistry where matching is often difficult, if not impossible, to achieve.
The longitudinal section of the study initially focused on the changes during the period of pre-surgical orthodontics. This is an area which has not been studied previously and the baseline for studying orthognathic patients had usually been taken as prior to surgery rather than prior to starting any treatment, albeit with no evidence that this was an appropriate assumption. It was therefore reassuring to find that the pre-surgical orthodontic phase appears to have minimal effect on the chosen psychometric outcome measures. Also of importance was the finding that the duration of pre-surgical orthodontics does not appear to have any adverse effects.

The section of the longitudinal study pertaining to T3/T2 and T3/T1 changes showed encouraging changes in the outcome variables. State anxiety, trait anxiety, self esteem, body image and facial body image showed significant improvements as a result of treatment. There were also a number of effects due to the explanatory variables, with gender showing the greatest differences. It must also be noted that variation in selection groups (i.e. unit, clinician, geographical area) may have had residual influences which were not accounted for.
6.0  CONCLUSIONS

Comparison of experimental and control groups

- This study showed that there were some differences in the psychological profile of orthognathic patients with respect to control group respondents.
- The orthognathic patient was shown to be more likely to display higher levels of state anxiety, greater social support (number), lower body image and lower facial body image. It is also possible that self esteem was lower in the experimental group, although statistical significance was only borderline at the 5% level. Whether these differences are clinically significant may form the basis for future research.

Comparison of questionnaires at T2/T1

- Pre-surgical orthodontics had minimal effect on the patient’s psychological profile and previous researchers were probably justified in using the pre-surgical measurement as a baseline measurement.
- The duration of pre-surgical orthodontics had no effect on the outcomes.
- Only body image showed significant change as a result of pre-surgical orthodontics.
- Older patients were more likely to show increases in state anxiety and may need to be monitored more carefully during treatment than younger age groups.
Comparison of questionnaires at T3/T2

- There were significant improvements due to treatment for several of the outcome variables: state anxiety; trait anxiety; self esteem; body image and facial body image.
- Gender was found to affect state and trait anxiety, with duration of pre-surgical orthodontics also having an effect. Age had a small effect on facial body image.

Comparison of questionnaires at T3/T1

- There were few significant explanatory variables and those which were observed were for body image and facial body image.
- Intercepts indicated significant improvements due to treatment for several of the outcome variables: state anxiety; trait anxiety; self esteem; body image and facial body image.

Statistical observations

- Multilevel modelling is a powerful statistical method that allows exploration of hierarchical dental data in a way which was not feasible previously using standard statistical techniques.
- To undertake this form of analysis effectively, one requires teaching in the use of appropriate statistical software and training in the interpretation of model findings. This highlights the importance of collaboration between clinicians and statisticians to utilise fully dental research data in the most appropriate and effective manner.
- Definite conclusions were sometimes difficult to draw by T3, due to the reduced number of respondents (n=42). This indicates that, for outcomes in this study, the number of subjects were around the threshold required to provide statistical validity for detailed analyses of this type, even when using multilevel multivariate
techniques. However, the threshold would be even higher if these techniques had not been employed.

- The outcome measures chosen were, on the whole, very informative. However, the social support scale was less reliable than initially hoped. It would seem that there is scope for some further development of instruments for this area of research.
8.0 SUGGESTIONS FOR FURTHER STUDY

The use of multilevel multivariate regression analysis could form the basis for analysis of data from many studies in the field of orthodontics.

In any study of this type it must always be borne in mind that the instruments which were chosen may not be sensitive enough to detect changes occurring as a result of orthognathic surgery. It may, therefore, be appropriate to repeat the study using different psychometric tests to confirm the findings from this study.
GENERAL CONCLUSIONS
GENERAL CONCLUSIONS

Research in the field of orthognathic surgery has traditionally focused on subjective outcomes. There remains a lack of research using outcome measures which are quantifiable, reliable, valid and responsive and this was the background to this thesis. Chapters 2 to 4 looked at three different areas of orthognathic treatment but the findings were consistent, with all three aspects showing positive changes associated with intervention. Chapter 2 showed that orthognathic treatment is an intervention which results in a good cost to benefit ratio and this type of information may prove useful if the provision of treatment under the NHS should become an issue. Chapter 3 looked at quality of life changes associated with treatment and the condition-specific measure which was developed shows promise for use in future clinical trials. Chapter 4 utilised a relatively new method of statistical modelling to look at changes in psychometric outcome measures, again with positive changes noted.

Overall, this research supports the view that orthognathic treatment is associated with positive effects for the patient and, in addition, the cost implications to the NHS are below the threshold set for funding. However, this is just the start of this area of research and there is now a need to establish larger multi-centre trials with sample sizes which are sufficiently large to affect policy making and possibly future funding decisions.

The first recommendation is, therefore, to develop a multi-centre trial for research into cost-utility of orthognathic treatment. This could also be combined with further validation of the OQLQ and will allow quality of life changes to be studied in greater depth.

The negative aspects associated with the process of treatment cannot be overlooked, however, and a number of findings from Chapter 4 merit further study. In particular, the effect of age and gender on some of the psychometric measures (i.e. anxiety). The negative aspects of treatment are of great importance to the patient and may influence satisfaction with their treatment as a whole. Therefore, a second suggestion for future
work would be to investigate the process of treatment, aiming to reduce negative aspects as much as possible. This may involve, for example: new methods of information provision; allocation of a liaison clinician (possibly a dental nurse) for each patient during treatment and increased involvement of previous patients in providing advice.

This form of research remains an important way forward in both medicine and dentistry and allows comparison of dental interventions with other health care outcomes. This is likely to be used increasingly in resource allocation in the future.
APPENDICES
APPENDIX 1

HEALTH STATE DESCRIPTIONS

Description of perfect health (based on EuroQol descriptions)

- Mobility: no problems walking about.
- Self-care: no problems with self-care.
- Activities: no problems with work; study; housework; family and social activities.
- Pain/Discomfort: no pain or discomfort.
- Anxiety/Depression: not anxious or depressed about anything.

Verbal description given to the general public group regarding a typical orthognathic patient (photographs shown in addition):

You will be:

- Physically healthy.
- Mobile and not restricted in physical activities.
- Self caring.
- Able to work.

However, you will be concerned about one or more of the following features:

- You may be concerned about/dislike your facial appearance.
- You may be concerned about/dislike the appearance of your teeth.
- You may get depressed about your appearance and sometimes feel that people are staring at you.
- You may feel self conscious in social situations and worry about meeting new people because of your appearance.
- You may have problems eating/biting/chewing certain foods.
## EXAMPLES OF STANDARD GAMBLE

<table>
<thead>
<tr>
<th>DENTOFACIAL DEFORMITY</th>
<th>GAMBLE</th>
<th>90% perfect health (utility = 0.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1.png" alt="Diagram" /></td>
<td><img src="image2.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DENTOFACIAL DEFORMITY</th>
<th>GAMBLE</th>
<th>10% perfect health (utility = 0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image3.png" alt="Diagram" /></td>
<td><img src="image4.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>

Following page:

**Flow diagram of a conventional Standard Gamble question**

(Reproduced from CHEPA Paper #90-9 by kind permission of Furlong *et al.*)
APPENDIX 1  UTILITY MEASURES

START

A

STEP 1

B = Inconsistent Utility >1.0

STEP 2

A: 10 vs 90

B = Indifferent Utility = 1.0

STEP 3

A: 90 vs 10

B = Indifferent Utility = 0.1

STEP 4

A: 20 vs 80

B = Indifferent Utility = 0.20

STEP 5

A: 80 vs 20

B = Utility = 0.85

STEP 6

A: 30 vs 70

B = Indifferent Utility = 0.80

STEP 7

A: 70 vs 30

B = Indifferent Utility = 0.70

STEP 8

A: 40 vs 60

B = Indifferent Utility = 0.40

STEP 9

A: 60 vs 40

B = Utility = 0.65

STEP 10

A: 50 vs 50

B = Utility = 0.55
EXAMPLES OF TIME TRADE-OFF

Dentofacial deformity for 50 years or:

Perfect health for 45 years then death

Perfect health for 5 years then death

Perfect health for 40 years then death

(utility = 0.95)

(utility = 0.05)

(utility = 0.85)
PATIENT A

- Rating scale

- Standard gamble

- Time trade-off
PATIENT B
- Rating scale

- Standard gamble

- Time trade-off
PATIENT C

- Rating scale

- Standard gamble

- Time trade-off
PATIENT D

- Rating scale

- Standard gamble

- Time trade-off
APPENDIX 3 RESOURCE USE

NB: All staff hourly rates include employer’s on-costs and London weighting.

ORTHODONTICS:

<table>
<thead>
<tr>
<th>OPERATOR TIME</th>
<th>Total treatment time based on the number of visits and the time allocated in the clinician’s diary for that patient (data from PAS system and operator’s diaries):</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Planning visits 30 minutes</td>
</tr>
<tr>
<td></td>
<td>• Bond-up 45 minutes</td>
</tr>
<tr>
<td></td>
<td>• Bands and AWs 45 minutes</td>
</tr>
<tr>
<td></td>
<td>• AW change 30 minutes, 15 or 30 minutes for Registrar</td>
</tr>
<tr>
<td></td>
<td>• Debond 45 minutes</td>
</tr>
<tr>
<td>Pay scales:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Registrar (mid grade) £14.74 per hour</td>
</tr>
<tr>
<td></td>
<td>• Senior Registrar (mid grade) £17.68 per hour</td>
</tr>
<tr>
<td></td>
<td>• Consultant (senior) £35.11 per hour</td>
</tr>
<tr>
<td></td>
<td>• Also 2.5 hours Consultant (or Senior Registrar) time during treatment for chairside supervision of registrars (based on 3 x 15 minute planning visits and 5 minutes each appointment) and 30 minutes for advice to SRs</td>
</tr>
<tr>
<td></td>
<td>• Casualty visits assumed to be 15 minutes each (Registrar and mid grade nurse)</td>
</tr>
<tr>
<td></td>
<td>• Also include 4 retainer check visits in the first year following treatment @ 15 minutes each</td>
</tr>
</tbody>
</table>

| NURSING COSTS | DA1 nurse @ £6-97 per hour                                                                      |
|              | Senior nurse @ £9-52 per hour                                                                    |
|              | (Assuming registrars work with a junior nurse and senior registrars and consultants work with a senior nurse) |

| STUDY MODELS | Start of treatment 2 sets                                                                      |
|              | Pre-surgery 1 set                                                                               |
|              | Debond 1 set                                                                                   |
|              | Cost per set of set of study models £6-00                                                       |
|              | (as charged by the lab these are sent to)                                                       |
|              | **Total per patient £24-00**                                                                    |

| RADIOGRAPHS | Based on EDH data:                                                                            |
|            | • Start of treatment - Lateral ceph; OPG; Ant occl                                              |
|            | • Pre-surgery - Lateral ceph; OPG                                                                 |
|            | • At wafer removal - Lateral ceph; OPG                                                          |
### APPENDIX 3 RESOURCE USE

- **Prior to debond - Lateral cep; OPG**
- **Follow-up - Lateral cep; OPG**

- Radiographer time for Lateral ceph and OPG = 15 minutes
- Radiographer time for Lateral ceph, OPG and anterior occlusal = 20 minutes
- Total radiographer time = 80 minutes

- Dark room technician time for developing 2 or 3 radiographs = 10 minutes
- Total = 50 minutes

- Materials cost (films/developing chemicals) = £2.00 per set of radiographs (£10.00 total)
- Radiographer pay: Senior 1 @ £14.10 per hour
- Dark room technician pay @ £6.69 per hour

**Total per patient £34.38**

<table>
<thead>
<tr>
<th>PHOTOGRAPHS/SLIDES</th>
<th>One set each at start of treatment; pre-surgery post-surgery and end of treatment = Total of 10 slides each time = 40 slides in total (approx. 2 films)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Also camera batteries etc.</td>
</tr>
<tr>
<td></td>
<td><strong>Total £25.00 per patient</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BRACKETS AND BANDS</th>
<th>Full arch bond up and 1&quot;/2&quot; molar bands banded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonds @£1-22 each</td>
<td>£24-40</td>
</tr>
<tr>
<td>Bands @0-98 each</td>
<td>£7.84</td>
</tr>
<tr>
<td><strong>Total £32.00 per patient</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ARCH WIRES</th>
<th>Upper and lower:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.14 NiTi</td>
<td>0.16 NiTi  0.018 NiTi  £24-40 for 10</td>
</tr>
<tr>
<td>0.16 steel</td>
<td>0.018 steel  £6-60 for 10</td>
</tr>
<tr>
<td>17 x 25 NiTi</td>
<td>19 x 25 NiTi  £32-68 for 10</td>
</tr>
<tr>
<td>18 x 25 steel</td>
<td>19x x25 steel  £7-70 for 10</td>
</tr>
<tr>
<td>17 x 25and 19 x 25 braided steel</td>
<td>£14-53 for 10</td>
</tr>
<tr>
<td><strong>Archwire costs calculated individually according to individual patient notes</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OTHER CONSUMABLES</th>
<th>Cost of gloves per patient (£5-20 per 50 pairs; £11-00 low allergy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cups and mouthwash tablets (cups £5-20 for 50)</td>
</tr>
<tr>
<td></td>
<td>Disposable 3 in 1 tips (£20 for 250 - assuming 30 used per patient)</td>
</tr>
<tr>
<td></td>
<td>Saliva ejectors (£1-79 per 100 - assuming 30 used)</td>
</tr>
<tr>
<td></td>
<td>Burs (£2-90 - new bur used for each debond)</td>
</tr>
<tr>
<td></td>
<td>Alginate (£4-52 per bag - assuming 1 bag used for 10 impressions)</td>
</tr>
<tr>
<td></td>
<td>Modules (assuming modules changed at every visit 20 visits x 20 modules)</td>
</tr>
<tr>
<td>Resource Use</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Ligatures</strong> (for surgery time)</td>
<td></td>
</tr>
<tr>
<td><strong>Elastics</strong> (assuming 4 bags pre-surgery and 6 post-op)</td>
<td></td>
</tr>
<tr>
<td><strong>Power chain</strong> (assuming one reel treats a minimum of 30 patients)</td>
<td></td>
</tr>
<tr>
<td><strong>NiTi closing coils</strong> (£3-20 each - assuming 4 per patient)</td>
<td></td>
</tr>
<tr>
<td><strong>Concise</strong> (£92-00 per pack - assuming one set bonds 50 patients)</td>
<td></td>
</tr>
<tr>
<td><strong>Etch</strong> (£6-88 per bottle - assuming one bottle treats 50 patients)</td>
<td></td>
</tr>
<tr>
<td><strong>Ketac Cement</strong> (£40-90 per set - assuming 30 sets of bands)</td>
<td></td>
</tr>
<tr>
<td><strong>Miscellaneous costs</strong> (mixing pad, brushes)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>£50-00 approx. per patient</td>
</tr>
</tbody>
</table>

**SURGICAL HOOKS**

- Mean number of hooks per patient = 8

- **Total** £15-00 per patient

**RETAINERS**

- Mean time taken to construct retainer (incl. pouring study models and construction of retainer) 2.5 hours

- **Hourly rates for technicians:**
  - MTO3 £11-36 per hour
  - MTO4 £15-10 per hour
  - MTO5 £18-71 per hour

- Assuming mid-grade technician at £15-10 per hour

- **Cost of materials:** maximum of £2-00 per retainer

- **Total** £39-75 per patient

**NB:** Additional appliances (RME; Quadhelix; Biteplanes) costed individually

**MISCELLANEOUS**

- **Receptionist** (A & C 3)
  - 1 hour at £8-42 per hour

- **Secretarial support** (A & C 4)
  - 0.5 hour at £9-30 per hour

- **Finding model boxes - receptionist** @ £8-42 per hour
  - 2 hours (3-4 minutes per visit)

- **Finding departmental notes - receptionist** @ £8-42 per hour
  - 2 hours (as above)

- **Finding hospital notes - medical records clerk** @ £8-42 per hour
  - 2 hours (as above)

- **Placement of study models in model boxes/ lab co-ordination - lab receptionist** @ £8-42/ hour
  - 0.5 hours

- **Total** £67-80 per patient

**Individual resource use documented in table at the end of this appendix**
## OUT-PATIENTS:

### JOINT CLINIC APPOINTMENTS

Mean duration of pre-surgery joint clinic appointment = 20 minutes (x2)

All patients have at least 2 appointments pre surgery; some have more if there is a question over the treatment plan.

Staff (costs as before) on each clinic:
- 1 x Orthodontic Consultant and 2 x SR
- 1 x Maxillofacial Consultant and 2 x SpR (£16-14 per hour)
- Orthodontic nurse and Maxfac nurse (1 senior and 1 mid grade)

= £154-35 per hour

In addition, 15 minutes with MaxFac SpR at start of treatment and pre-surgery to provide information re surgery (30 minutes for each patient) = £8-07

**Total: £110-97 based on two 20 minute appointments (additional visits costed as appropriate)**

### WAFER IMPRESSIONS AND TRY-IN

One appointment for impressions and facebow registration
(Maxfac Senior House Officer £14-95 per hour)

- Duration of appointment 1.5 hours
- Wafer try in 30 mins
- Materials £2-00

**Total: £31-90**

### SURGICAL WORK-UP AND WAFER CONSTRUCTION

Mean duration: Approx 6.5 hours (including time on the clinic checking work-up, casting models, articulation, model surgery and wafers)

Technician hourly rate MTO4 @ £15-10 per hour

£98-15

Cost of materials £10-00

**Total: £108-15**

### PRE-OP CLERKING

Mean duration 2 hours (incl. checking for blood results etc.)

SHO hourly rate £14.95

**Total £29-90**

### BLOOD TESTS

- Full blood count (pre- and post-op) £5-00
- Urea and electrolytes (post-op) £4-75
- Cost of grouping and saving each patient (not UCLH prices. Based on known prices for another trust) £12-00

**Total £21-75**
### POST-OP APPOINTMENTS

Average time of visits on the main clinic with all personnel present:
- one week post-op: 10 minutes
- two weeks post-op: 10 minutes
- also, 20 minutes with 1 x Ortho SR and 1 x MaxFac SpR for wafer removal, placement of elastics, post-op radiographs etc.
- Additional appointments if problems are experienced (some additional appointments on an individual consultant clinic rather than Joint clinic)
- Any additional clinics costed individually

### FOLLOW-UP APPOINTMENTS

@ 1 year; 2 years; 5 years and 10 years
- All reviews seen on the main clinic
- Average duration of appointments = 10 mins (total 40 mins) (costings as before)

Total: £102-90

### MISCELLANEOUS

- Receptionist time
  0.5 hour at £8-42 per hour
- Secretarial time
  0.5 hour at £9-30 per hour
- Finding model boxes: receptionist @ £8-42 per hour 0.5 hours
- Finding receptionist @ £8-42 per hour 1 hour

Total: £21-49

Individual resource use documented in table at the end of this appendix

### IN-PATIENT/HOTEL COSTS:

<table>
<thead>
<tr>
<th>IN-PATIENT STAY</th>
<th>£150-00 per night including nursing, drugs, consumables, cleaning, heat/ light</th>
</tr>
</thead>
</table>

(NB: Patients return directly to the ward, no high dependency unit costs)

Individual resource use documented in table at the end of this appendix
MAIN THEATRE COSTS:

- The number of orthognathic cases per day is restricted to two due to bed availability. Bimaxillary procedure are booked for one session and a single jaw procedure for 2/3 of a session (including anaesthetic time, transfers etc.).
- Surgical staff and anaesthetic staff costed according to those personnel recorded on individual operating notes.
- Additional surgical procedures (e.g. rhinoplasty, removal of plates) costed based on time required for individual procedures.

<table>
<thead>
<tr>
<th>METHOD 1</th>
<th></th>
</tr>
</thead>
</table>
| **THEATRE OVERHEADS** | • £0.96/ minute  
• UCLH/ Middlesex Hospital costings  
• NOT including any clinical staff, consumables  
• 4 hour sessions |
| **CLINICAL STAFF COSTS** | • Consultant (Senior) £35.11 per hour  
• SpR (Mid scale) £16.14 per hour  
• SHO (Mid scale) £14.95 per hour  
  Surgeons costed according to personnel listed in theatre notes  
  Anaesthetists: Consultant and SpR routinely present |
| **NURSING STAFF** | • Based on mid F Grade staff  
• 2 nursing staff in theatre (1 scrub nurse and 1 runner) @ £12-60/ hour  
• 1 in recovery: based on mid E grade @ £11-27/ hour  
  (assuming patient spends 1 hour in recovery)  
  (Data provided by Chelsea and Westminster hospital based on National Pay structure for nurses) |
| **OTHER OPERATING DEPARTMENT STAFF** | One ODA: mid E grade @ £11-27/ hour |
| **CONSUMABLES** | General consumables:  
- Antiembolism (TED) stockings - 1 pair £3-66  
- 5 x sterile gowns @ £2-68 each (surgical staff only) £13-40  
- 5 pairs of sterile gloves @ £63-95/50 £6-40  
- 7 x hats @ £10-62/100 (incl. anaesthetist and ODA) £0-74  
- 7 x visor type masks @ £27.64/25 £7-74  
- 5 x scrubbing brushes @ £11.71/30 £1-95  
- Light handles (per pair) £1-12 |
APPENDIX 3 RESOURCE USE

**Betadine preparation (10% solution)** £1-91
**Yankeur sucker** £0-54
**Sucker tubing** £1-35
**Suction liner** £2-16
**Drapes (large £4-17 and 2 x small £0-83 each)** £5-83
**Op-site drape @ £43-30/10** £4-33

**Total £51-13**

**Oral and Maxillofacial surgery consumables:**
- **Saw blade @ £188.00/5** £37-60
- **10 4 inch swabs (X-ray detectable) @ £209-90/1000** £2-10
- **3 x No 15 scalpel blades @ £6-75/100** £0-20
- **5 x Blue gauze swabs @ £4-29/50** £0-43
- **2 x Tonsil swabs @ £210-02/400** £1-05
- **Diathermy pad @ £78-25/100** £0-78
- **Diathermy tip @ £91-00/50** £1-82
- **Discard-A-Pad @ £21-43/50** £0-43
- **3 x 20ml syringes @ £4-39/50** £0-26
- **2 x pink needles @ £3-69/100** £0-07
- **1 x green needles @ £1-55/100** £0-02
- **2 x kidney dish** £0-66
- **2 x gallipot @ £1-49/25** £0-06
- **2 x 3.0 Vicryl sutures @ £19-12/12** £3-19
- **1 x 5.0 proline suture @ £48-87/24** £2-04

**Total £50-71**

**Anaesthetic consumables:**
- **14 gauge Abbocath** £0-70
- **20 gauge Abbocath** £0-70
- **Syringes (based on 1 x 20ml and 10 x 5 or 10ml)** £0-63
- **Arterial pressure monitoring set** £6-57
- **3 x ECG dots @ £28-60/600** £0-14
- **Catheter mount @ £19-15/25** £0-77
- **Portex nasal tube with ivory soft seal cuff @ £43-71/10** £4-37
- **Paraffin coated throat pack @ £13-67/50** £0-27
- **Filter with bacterial/viral sampling port** £1-16
- **2 x Infusion/giving set** £2-42
- **Y connector for IV infusion @ £67-56/50** £1-35
- **Warming blanket (Warm touch system)** £15-20
- **Disposable temperature probe @ £65-99/1000** £0-07
- **Micropore tape** £0-86

**Total £35-21**

**Total: £137-05**

**DRUGS USED (BASED ON BNF PRICES FOR YEAR 2000 AND UCLH PHARMACY PRICES)**

**General:**
- **Temazapam 20mg (premed)** £0-06
- **Ephedrine nose drops (premed)** £1-15
- **Chloramphenicol eye ointment** £1-10
- **Hydrocortisone ointment** £0-38
- **Fragmin 2500 units/ sc pre-op (anticoagulant)** £1-86
- **Marcain and adrenaline (op site analgesia)** £1-23

**Total £5-78**
### Anaesthetic drugs:
- Propofol induction (200mg) £3-90
- N₂O / O₂ / Desflurane maintenance (difficult to cost as gases are supplied centrally to the whole hospital, estimate cost) £10-00
- Desflurane (approx. one third to half of bottle) max. £25-00
- 2g Tranexamic acid IV (haemostatic) £5-16
- Alfentanil 10ml (enhancement of anaesthesia) £3-31
- Vecuronium 10mg (muscle relaxant) £4-44
- Clonidine 150μg (antihypertensive) £0-30
- Voltarol (analgesia) 100mg/pr x2 £0-68
- Dexamethasone 8mg IV on induction + 3 doses after £7-04
- Amoxicillin 1g IV on induction + 3 x 500mg dose after £3-29

**Total £63-12**

### Variable administration of:
- Cyclizine IV (anti-emetic) £0-54
- Morphine 10mg IV/IM x 2 doses (analgesia) £1-32

**Total £1-86**

### Fluid requirements:
- Fluid (typical example: 2litre Hartmann’s) £3-00
- 1 litre Gelafusine £6-80
- Saline flush (several) £0-50

**Total £10-30**

### Drugs provided on the ward:
- Covered by *per diem* cost per night

### Drugs to take home at discharge:
- Corsodyl mouthwash £1-93
- Co-proxamol analgesia (40) £0-60
- Amoxycillin for 5 days (500mg x 20) £2-21

**Total £4-74**

**Total: £85-80**

### COSTS OF PLATES; BURS

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set of burs</td>
<td>£17-64</td>
</tr>
<tr>
<td>Lindeman burs</td>
<td>£15-31</td>
</tr>
<tr>
<td>4 x mid “L” plates (maxilla)</td>
<td>£165-57</td>
</tr>
<tr>
<td>16 x 7mm screws (maxilla)</td>
<td>£125-57</td>
</tr>
<tr>
<td>6 x 13mm screws (mandible)</td>
<td>£55-38</td>
</tr>
<tr>
<td>Additional screws as required</td>
<td></td>
</tr>
</tbody>
</table>

**Total**
- Bimaxillary Surgery £379-47
- Mandibular surgery only £88-33
- Le Fort 1 only £324-09

### WEAR AND TEAR OF INSTRUMENTS; CCSD COSTS (incl. staff)

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maxfac 1 set; Maxfac 2 set; Hall air drill ; King Combo bone plating set ; Bone hook (to reprocess); Bipolar forceps and leads</td>
<td>£77-72</td>
</tr>
</tbody>
</table>

(as calculated by UCLH theatre and CCSD staff)

353
<table>
<thead>
<tr>
<th>METHOD 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>THEATRE COST PER SESSION</td>
<td>£1947-00 per 4 hour session (including nursing staff, anaesthetist, drugs, consumables, lighting etc.)</td>
</tr>
<tr>
<td>CLINICAL STAFF</td>
<td>As in Method 1</td>
</tr>
<tr>
<td>NURSING STAFF</td>
<td>Included</td>
</tr>
<tr>
<td>COST OF PLATES, BURS ETC.</td>
<td>As in Method 1</td>
</tr>
<tr>
<td>WEAR AND TEAR/CCSD COSTS</td>
<td>Included</td>
</tr>
</tbody>
</table>

Individual resource use documented in table at the end of this appendix
### DAY STAY THEATRE

(Removal of infected plates; closure of palatal fistula; wound exploration)

(NB: One patient had infected plates/graft removed in main theatre)

<table>
<thead>
<tr>
<th>METHOD 1</th>
<th></th>
</tr>
</thead>
</table>
| **THEATRE OVERHEADS** | • £0.96/ minute  
  • NOT including any clinical staff, consumables |
| **CLINICAL STAFF COSTS** | • Consultant (senior grade)  
  £35.11 per hour  
  • SpR (mid-grade)  
  £16.14 per hour  
  • SHO (mid-grade)  
  £14.95 per hour  

Surgeons - according to those listed on theatre notes  
Anaesthetist - Consultant routinely present |
| **NURSING STAFF** | Theatre  
  • 2 Grade E nursing staff in theatre (1 scrub nurse and 1 runner): both at £9-76/ hour  
  • 1 Grade E nurse in recovery @ £9-76/ hour (assuming patient spends 1 hour in recovery)  

Ward staff  
  • 2 Grade E nurses (one senior) @ £9-76 and £10-47/ hour  
  • Assume a maximum of 4 patients for two nurses spending 6/7 hours in total = approx. £40-00 per patient in ward nursing costs |
| **OPERATING DEPARTMENT STAFF** | One ODA: Maximum grade is E @ £10-47 for maximum grade |
| **CONSUMABLES** | General consumables:  
  • 4 x sterile gowns @ £2-68 each  
  £10-72  
  • 4 pairs of sterile gloves @ £63-95/50  
  £5-12  
  • 6 x hats @ £10-62/100  
  £0-64  
  • 6 x visor type masks @ £27.64/25  
  £6-63  
  • 4 x scrubbing brushes @ £11.71/30  
  £1-56  
  • Light handles (per pair)  
  £1-12  
  • Betadine preparation (10% solution)  
  £1-91  
  • Yankeur sucker  
  £0-54  
  • Sucker tubing  
  £1-35  
  • Suction liner  
  £2-16  
  • Drapes (large £4-17 and 2 x small £0-83 each)  
  £5-83 |
### APPENDIX 3 RESOURCE USE

#### Oral and Maxillofacial surgery consumables:
- **5 x 4 inch swabs (X-ray detectable)** @ £209-90/1000  
  £1-05
- **1 x No 15 scalpel blades** @ £6-75/100  
  £0-07
- **5 x Blue gauze swabs** @ £4.29/ 50  
  £0-43
- **2 x Tonsil swabs** @ £210-02/400  
  £1-05
- **Diathermy pad** @ £78-25/100  
  £0-78
- **Diathermy tip** @ £91-00/50  
  £1-82
- **Discard-A-Pad** @ £21-43/50  
  £0-43
- **1 x 20ml syringes** @ £4-39/50  
  £0-09
- **1 x pink needles** @ £3.69/100  
  £0-04
- **1 x green needles** @ £1-55/100  
  £0-02
- **2 x kidney dish**  
  £0-66
- **2 x gallipot** @ £1-49/25  
  £0-06
- **1 x 3.0 Vicryl sutures** @ £19-12/ 12  
  £1-59
- **1 x 5.0 proline suture** @ £48-87/ 24  
  £2-04

**Total £41-91**

#### Anaesthetic consumables:
- **14 gauge Abbocath Syringes**  
  £0-70
- **3 x ECG dots** @ £28-60/600  
  £0-14
- **Catheter mount** @ £19-15/25  
  £0-77
- **Portex nasal tube with ivory soft seal cuff** @ £43-71/10  
  £4-37
- **Paraffin coated throat pack** @ £13-67/50  
  £0-27
- **Filter with bacterial/ viral sampling port**  
  £1-16
- **1 x Infusion/ giving set**  
  £1-21
- **Micropore tape**  
  £0-86

**Total £10-13**

**Total: £61-97**

#### DRUGS USED

(BASED ON UCLH PHARMACY PRICES AND BNF PRICES FOR YEAR 2000)

**General:**
- **Chloramphenicol eye ointment**  
  £1-10
- **Hydrocortisone ointment**  
  £0-38
- **Marcain and adrenaline (op site analgesia)**  
  £1-23

**Total £2-71**

**Anaesthetic drugs:**
- **Propofol induction (200mg)**  
  £3-90
- **N₂O/ O₂ / Isoflurane maintenance (gases)**  
  £10-00
- **Isoflurane** (approx. one third to half of bottle) max.  
  £15-00
- **2g Tranexamic acid IV (haemostatic)**  
  £5-16
- **Alfentanil 10ml** (enhancement of anaesthesia)  
  £3-31
- **Vecuronium 10mg** (muscle relaxant)  
  £4-44
- **Clonidine 150µg** (antihypertensive)  
  £0-30
- **Voltarol** (analgesia) 100mg/pr x2  
  £0-68
- **Dexamethasone 8mg IV on induction**  
  £1-76
- **Amoxicillin 1g IV on induction + 1 x 500mg dose after**  
  £1-97

**Total £41-52**
**Drugs to take home at discharge:**
- Corsodyl mouthwash £1.93
- Co-proxamol analgesia (40) £0.60
- Amoxycillin for 5 days (500mg x 20) £2.21

Total £4.74

**Total: £48.97**

<table>
<thead>
<tr>
<th>WEAR AND TEAR OF INSTRUMENTS; CCSD COSTS (incl. staff)</th>
<th>Oral surgery 1 set (including screw driver)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total £25.00</td>
</tr>
</tbody>
</table>

**DAY STAY THEATRE COSTING**

<table>
<thead>
<tr>
<th>THEATRE COST PER SESSION</th>
<th>£750.00/session (based on a 4 hour session)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Includes theatre nurses; anaesthetist; consumables administration support</td>
</tr>
</tbody>
</table>

**SURGEONS**

- As costed above

**WARD COSTS**

- 2 x Grade E nurses
- Based on 4 of patients on the ward for 6/7 hours

Individual resource use documented in table at the end of this appendix
## MISCELLANEOUS:

<table>
<thead>
<tr>
<th>HYGIENIST</th>
<th>£10-97 per hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXTRactions</td>
<td>Based on cost per visit for dental specialties of £42-00 (assuming 1 visit for extraction of 1 or 2 teeth and 2 visits for extraction of 3 or 4 teeth)</td>
</tr>
</tbody>
</table>
| ADVANCED RESTORATIVE CARE (if required) | Implant costs - Based on Nobelbiocare prices for the Bränemark system (excl. VAT):
  **Surgical components:**
  - Drill kit £13-00
  - Twist drill 3mm £11-00
  - Countersink £13-00
  - Titanium fixture (each) £164-00
  - Cover screw (each) £26-00
  - Healing abutment (each) £26-00
  **Prosthetic components:**
  - Estheticone™ abutment (each) £102-50
  - Healing cap (each) £4-00
  - Temporary cylinder for Estheticone (each) £15-00
  - Gold cylinder, conical with chamfer (each) £47-00
  - Gold screw (each) £23-00
  - Abutment replica, conical (each) £5-50
  - Clinician time for 2 units is 4 hours; for 3 units is 5 hours (Consultant)
  - Technician time for 2 unit fixture is 5 hours; for 3 units is 6 hours (senior technician @ £15-10/hour)
  - Cost for gold/ porcelain £50-00 |
| MISCELLANEOUS      | Receptionist time 0.5 hour at £8-42 per hour
  Secretarial time 0.5 hour at £9-30 per hour
  Finding hospital notes - medical records clerk 0.5 hours @ £8-42 per hour |

Total: £13-07

Individual resource use documented in table at the end of this appendix
## INDIVIDUAL PATIENT DATA

<table>
<thead>
<tr>
<th>NUMBER</th>
<th>ORTHODONTIC TREATMENT UNDERTAKEN BY</th>
<th>MAXILLOFACIAL PERSONNEL INVOLVED IN SURGERY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td>2</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 2 x SpR</td>
</tr>
<tr>
<td>3</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td></td>
<td>(Repeat bimax: 1 x Consultant and 2 x SpR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Exploration of wound: 1 x Consultant; 1 x SpR)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td>5</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td></td>
<td>(Exploration of wound: 1 x SpR; 1 x SHO)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td>7</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td>8</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 1 x SHO</td>
</tr>
<tr>
<td>9</td>
<td>Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td></td>
<td>(Removal of plates: 1 x SpR; 1 x SHO)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 3 x SpR</td>
</tr>
<tr>
<td>11</td>
<td>Senior Registrar/Registrar</td>
<td>1 x Consultant; 2 x SpR; 1 x SHO</td>
</tr>
<tr>
<td></td>
<td>(Rhinoplasty: 1 x Consultant; 2 x SpR)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 2 x SpR</td>
</tr>
<tr>
<td>13</td>
<td>Senior Registrar</td>
<td>1 x Consultant; 2 x SpR</td>
</tr>
<tr>
<td>14</td>
<td>Registrar</td>
<td>1 x Consultant; 2 x SpR</td>
</tr>
<tr>
<td>15</td>
<td>Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td>16</td>
<td>Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td>17</td>
<td>Senior Registrar/Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td></td>
<td>Registrar</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>18</td>
<td>Registrar</td>
<td>1 x Consultant; 2 x SpR</td>
</tr>
<tr>
<td>19</td>
<td>Registrar/ Senior Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
<tr>
<td>20</td>
<td>Registrar</td>
<td>1 x Consultant; 2 x SpR</td>
</tr>
<tr>
<td>21</td>
<td>Registrar</td>
<td>1 x Consultant; 1 x SpR; 1 x SHO</td>
</tr>
</tbody>
</table>
## INDIVIDUAL RESOURCE USE (all values in £ sterling)

<table>
<thead>
<tr>
<th>ORTHODONTICS (incl. hygienist and casualty visits)</th>
<th>EXTRACTIONS</th>
<th>RESTORATIVE WORK</th>
<th>OUT-PATIENT VISITS</th>
<th>IN-PATIENT COSTS</th>
<th>THEATRE COSTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASE ESTIMATE</td>
<td></td>
<td></td>
<td></td>
<td>BASE EST.</td>
<td>METHOD 1</td>
</tr>
<tr>
<td>1</td>
<td>£701.58</td>
<td>£0</td>
<td>£0</td>
<td>£750.00</td>
<td>£1,537.82</td>
</tr>
<tr>
<td>2</td>
<td>£741.87</td>
<td>£0</td>
<td>£0</td>
<td>£300.00</td>
<td>£967.83</td>
</tr>
<tr>
<td>3</td>
<td>£770.43</td>
<td>£0</td>
<td>£0</td>
<td>£600.00</td>
<td>£2,775.99</td>
</tr>
<tr>
<td>4</td>
<td>£658.33</td>
<td>£42.00</td>
<td>£0</td>
<td>£600.00</td>
<td>£1,537.82</td>
</tr>
<tr>
<td>5</td>
<td>£584.39</td>
<td>£0</td>
<td>£0</td>
<td>£600.00</td>
<td>£1,876.31</td>
</tr>
<tr>
<td>6</td>
<td>£734.59</td>
<td>£0</td>
<td>£0</td>
<td>£600.00</td>
<td>£1,537.82</td>
</tr>
<tr>
<td>7</td>
<td>£667.66</td>
<td>£0</td>
<td>£0</td>
<td>£450.00</td>
<td>£1,537.82</td>
</tr>
<tr>
<td></td>
<td>£656.18</td>
<td>£0</td>
<td>£0</td>
<td>£542.17</td>
<td>£450.00</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>------</td>
<td>------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>9</td>
<td>£676.49</td>
<td>£42.00</td>
<td>£0</td>
<td>£561.08</td>
<td>£750.00</td>
</tr>
<tr>
<td></td>
<td>(Includes extra procedure to remove infected graft and plates)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>£681.54</td>
<td>£84.00</td>
<td>£0</td>
<td>£490.72</td>
<td>£450.00</td>
</tr>
<tr>
<td></td>
<td>(Includes one night for extra procedure)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>£688.56</td>
<td>£0</td>
<td>£1,704.86</td>
<td>£567.90</td>
<td>£750.00</td>
</tr>
<tr>
<td></td>
<td>(3 implants and superstructure)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Includes one night for rhinoplasty)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>£708.50</td>
<td>£42.00</td>
<td>£0</td>
<td>£567.90</td>
<td>£450.00</td>
</tr>
<tr>
<td>13</td>
<td>£732.98</td>
<td>£0</td>
<td>£0</td>
<td>£490.72</td>
<td>£450.00</td>
</tr>
<tr>
<td>14</td>
<td>£727.02</td>
<td>£0</td>
<td>£0</td>
<td>£722.25</td>
<td>£750.00</td>
</tr>
<tr>
<td>15</td>
<td>£761.80</td>
<td>£0</td>
<td>£0</td>
<td>£490.72</td>
<td>£450.00</td>
</tr>
<tr>
<td>16</td>
<td>£806.80</td>
<td>£0</td>
<td>£0</td>
<td>£516.45</td>
<td>£450.00</td>
</tr>
<tr>
<td>17</td>
<td>£738.31</td>
<td>£0</td>
<td>£0</td>
<td>£593.62</td>
<td>£450.00</td>
</tr>
</tbody>
</table>

362
<table>
<thead>
<tr>
<th>No.</th>
<th>Description</th>
<th>First Field</th>
<th>Second Field</th>
<th>Third Field</th>
<th>Fourth Field</th>
<th>Fifth Field</th>
<th>Sixth Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>£820.55 (Includes referral to Oral Medicine and treatment for angular cheilitis)</td>
<td>£84.00</td>
<td>£0</td>
<td>£567.90</td>
<td>£450.00</td>
<td>£967.83</td>
<td>£1,566.04</td>
</tr>
<tr>
<td>19</td>
<td>£639.74</td>
<td>£42.00</td>
<td>£1,296.73</td>
<td>£542.17</td>
<td>£300.00</td>
<td>£964.65</td>
<td>£1,562.86</td>
</tr>
<tr>
<td>20</td>
<td>£655.20</td>
<td>£0</td>
<td>£0</td>
<td>£553.44</td>
<td>£450.00</td>
<td>£1,901.23</td>
<td>£2,869.78</td>
</tr>
<tr>
<td></td>
<td>(Includes extra procedure for closure of palatal fistula)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>£715.23</td>
<td>£0</td>
<td>£0</td>
<td>£542.17</td>
<td>£450.00</td>
<td>£1,537.82</td>
<td>£2,591.27</td>
</tr>
</tbody>
</table>
Please read the following statements carefully. Please circle N/A or 1, 2, 3, 4 where:

- **N/A** means the issue covered by the statement either does not apply to you or if it does apply to you, it does not bother you at all.
- 1 means the issue covered in the statement bothers you a little.
- 4 means the issue covered in the statement bothers you a lot.
- 2 + 3 lie between “a little” and “a lot”.

<table>
<thead>
<tr>
<th>1. I am self-conscious about the appearance of my teeth</th>
<th>2. have problems biting</th>
<th>3. I have problems chewing</th>
<th>4. There are some foods I avoid eating because the way my teeth meet makes it difficult</th>
<th>5. I don’t like eating in public places</th>
<th>6. I get pains in my face or jaw</th>
<th>7. I don’t like seeing a side view of my face (profile)</th>
<th>8. I spend a lot of time studying my face in the mirror</th>
<th>9. I spend a lot of time studying my teeth in the mirror</th>
<th>10. I dislike having my photograph taken</th>
<th>11. I dislike being seen on video</th>
<th>12. I often stare at other people’s teeth</th>
<th>13. I often stare at other people’s faces</th>
<th>14. I am self-conscious about my facial appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Item Description</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. I try to cover my mouth when I meet people for the first time</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. I worry about meeting people for the first time</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. I worry that people will make hurtful comments about my appearance</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. I lack confidence when I am out socially</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. I do not like smiling when I meet people</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. I sometimes get depressed about my appearance</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. I sometimes think that people are staring at me</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Comments about my appearance really upset me, even when I know people are only joking</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Scoring of OQLQ**

Addition of individual item scores within each domain
Possible distribution of OQLQ for the future

Please read the following statements carefully. Please circle N/A or 1, 2, 3, 4 where:

N/A means the issue covered by the statement either does not apply to you or if it does apply to you, it does not bother you at all

1 means the issue covered in the statement bothers you a little

4 means the issue covered in the statement bothers you a lot

2 + 3 lie between “a little and a lot”

<table>
<thead>
<tr>
<th>1. I am self-conscious about the appearance of my teeth</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. I don’t like seeing a side view of my face (profile)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3. I dislike having my photograph taken</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4. I dislike being seen on video</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5. I am self-conscious about my facial appearance</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6. I try to cover my mouth when I meet people for the first time</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7. I worry about meeting people for the first time</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8. I worry that people will make hurtful comments about my appearance</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>9. I lack confidence when I am out socially</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10. I do not like smiling when I meet people</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>11. I sometimes get depressed about my appearance</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>12. I sometimes think that people are staring at me</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>13. Comments about my appearance really upset me, even when I know people are only joking</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Question</td>
<td>Options</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. have problems biting</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. I have problems chewing</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. There are some foods I avoid eating because the way my teeth meet makes it difficult</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. I don’t like eating in public places</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. I get pains in my face or jaw</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. I spend a lot of time studying my face in the mirror</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. I spend a lot of time studying my teeth in the mirror</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. I often stare at other people’s teeth</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. I often stare at other people’s faces</td>
<td>1 2 3 4</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 5  SHORT FORM-36

This series of questions asks about your views about your health. Some of the questions may seem irrelevant to you but it is important that you answer each question as indicated.

1. In general, would you say your health is:
   1. Excellent
   2. Very good
   3. Good
   4. Fair
   5. Poor

2. Compared to one year ago, how would you rate your health in general now?
   1. Much better than 1 year ago
   2. Somewhat better than 1 year ago
   3. About the same as 1 year ago
   4. Somewhat worse than 1 year ago
   5. Much worse than 1 year ago

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>YES, LIMITED A LOT</th>
<th>YES, LIMITED A LITTLE</th>
<th>NO, NOT LIMITED AT ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Vigorous activities such as running, lifting heavy objects, participating in strenuous sports</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>b. Moderate activities such as moving a table, pushing a vacuum cleaner, bowling or playing golf</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>c. Lifting or carrying groceries</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>d. Climbing several flights of stairs</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>e. Climbing one flight of stairs</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>f. Bending, kneeling or stooping</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>g. Walking more than a mile</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>h. Walking half a mile</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>i. Walking one hundred yards</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>j. Bathing or dressing yourself</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

<table>
<thead>
<tr>
<th>(circle one number on each line)</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down on the amount of time you spend on work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b. Accomplished less than you would like</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c. Were limited in the kind of work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d. Had difficulty performing the work or other activities (for example, it took extra effort)</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

5. During the past 4 weeks have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

<table>
<thead>
<tr>
<th>(circle one number on each line)</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Cut down the amount of time you spent on work or other activities</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b. Accomplished less than you would like</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c. Didn’t do work or other activities as carefully as usual</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours or groups?

1. Not at all
2. Slightly
3. Moderately
4. Quite a bit
5. Extremely
7. How much **bodily** pain have you had during the **past 4 weeks**?
   1. None
   2. Very mild
   3. Mild
   4. Moderate
   5. Severe
   6. Very severe

8. During the **past 4 weeks**, how much did pain interfere with your normal work (including both work outside the home and housework)?
   1. Not at all
   2. A little bit
   3. Moderately
   4. Quite a bit
   5. Extremely

9. For each question, please give one answer that comes closest to the way you have been feeling.
   How much of the time during the **past 4 weeks**:
   (circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>All of the time</th>
<th>Most of the time</th>
<th>A good bit of the time</th>
<th>Some of the time</th>
<th>A little bit of the time</th>
<th>None of the time</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Did you feel full of life?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>b. Have you been a very nervous person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>c. Have you felt so down in the dumps that nothing could cheer you up?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>d. Have you felt calm and peaceful?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>e. Did you have a lot of energy?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>f. Have you felt downhearted and low?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>g. Did you feel worn out?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>h. Have you been a happy person?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>I. Did you feel tired?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
10. During the **past 4 weeks**, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends or relatives, going out etc.)?

1. All of the time  
2. Most of the time  
3. Some of the time  
4. A little of the time  
5. None of the time

11. How **TRUE** or **FALSE** is each of the following statements for you?

   (circle one number on each line)

<table>
<thead>
<tr>
<th></th>
<th>Definitely true</th>
<th>Mostly true</th>
<th>Don’t know</th>
<th>Mostly false</th>
<th>Definitely false</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. I seem to get ill more easily than other people</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>b. I am as healthy as anybody I know</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>c. I expect my health to get worse</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d. My health is excellent</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
### SCORING FOR SF-36

#### PHYSICAL FUNCTIONING

<table>
<thead>
<tr>
<th>3a - j</th>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, limited a lot</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Yes, limited a little</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>No, not limited at all</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

#### ROLE - PHYSICAL

<table>
<thead>
<tr>
<th>4a - d</th>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

#### BODILY PAIN

<table>
<thead>
<tr>
<th>7 and 8</th>
<th>Item 7</th>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 7</td>
<td></td>
<td>None</td>
<td>1</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Very mild</td>
<td>2</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mild</td>
<td>3</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moderate</td>
<td>4</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Severe</td>
<td>5</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Very severe</td>
<td>6</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Item 8 (if 7 is answered)**

<table>
<thead>
<tr>
<th>Precoded score and Item 7</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>6</td>
</tr>
<tr>
<td>Not at all</td>
<td>5</td>
</tr>
<tr>
<td>A little bit</td>
<td>4</td>
</tr>
<tr>
<td>Moderately</td>
<td>3</td>
</tr>
<tr>
<td>Quite a bit</td>
<td>2</td>
</tr>
<tr>
<td>Extremely</td>
<td>1</td>
</tr>
</tbody>
</table>

#### GENERAL HEALTH

<table>
<thead>
<tr>
<th>1; 11a, b, c, d</th>
<th>ITEM 1</th>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>1</td>
<td>5.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>2</td>
<td>4.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>3</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>4</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>5</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### APPENDIX 5 SHORT FORM-36 QUESTIONNAIRE

#### ITEMS 11a and 11c

<table>
<thead>
<tr>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely true</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mostly true</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Don't know</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Mostly false</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Definitely false</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

#### ITEMS 11b and 11d

<table>
<thead>
<tr>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitely true</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Mostly true</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Don't know</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Mostly false</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Definitely false</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

#### VITALITY

**Items 9a, e, g, i**

#### ITEMS 9a and 9e

<table>
<thead>
<tr>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>All of the time</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Most of the time</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>A good bit of the time</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Some of the time</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>A little of the time</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>None of the time</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

#### ITEMS 9g and 9i

<table>
<thead>
<tr>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>All of the time</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Most of the time</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>A good bit of the time</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Some of the time</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>A little of the time</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>None of the time</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

#### SOCIAL FUNCTIONING

**Q6 and 10**

**Q6**

<table>
<thead>
<tr>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Slightly</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Moderately</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Quite a bit</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Extremely</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>
### APPENDIX 5  SHORT FORM-36 QUESTIONNAIRE

<table>
<thead>
<tr>
<th>Q10</th>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All of the time</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Most of the time</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Some of the time</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>A little of the time</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>None of the time</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

**ROLE - EMOTIONAL**

<table>
<thead>
<tr>
<th>Q5a - 5c</th>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

**MENTAL HEALTH**

<table>
<thead>
<tr>
<th>Q9 b, c, d, f, h</th>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All of the time</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Most of the time</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A good bit of the time</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Some of the time</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>A little of the time</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>None of the time</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q9 d, &amp; 9h</th>
<th>Response choice</th>
<th>Precoded score</th>
<th>Final score</th>
</tr>
</thead>
<tbody>
<tr>
<td>All of the time</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Most of the time</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>A good bit of the time</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Some of the time</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>A little of the time</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>None of the time</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**REPORTED HEALTH TRANSITION**

Precoded and final values are the same
### Transformed Score

<table>
<thead>
<tr>
<th>Scale</th>
<th>Lowest and Highest Raw Scores</th>
<th>Possible Raw Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>10, 30</td>
<td>20</td>
</tr>
<tr>
<td>R-P</td>
<td>4, 8</td>
<td>4</td>
</tr>
<tr>
<td>BP</td>
<td>2, 12</td>
<td>10</td>
</tr>
<tr>
<td>GH</td>
<td>5, 25</td>
<td>20</td>
</tr>
<tr>
<td>VT</td>
<td>4, 24</td>
<td>20</td>
</tr>
<tr>
<td>SF</td>
<td>2, 10</td>
<td>8</td>
</tr>
<tr>
<td>R-E</td>
<td>3, 6</td>
<td>3</td>
</tr>
<tr>
<td>MH</td>
<td>5, 50</td>
<td>25</td>
</tr>
</tbody>
</table>

Transformed scale = \( \frac{\text{Actual raw score} - \text{least possible raw score}}{\text{Possible raw score range}} \times 100 \)
**CONVERSION PROGRAMME**

to convert the eight components to two components (MCS and PCS)

Compute \( aPF = \frac{PF - 84.52404}{22.8949} \)
Compute \( aRP = \frac{RP - 81.19907}{33.79729} \)
Compute \( aBP = \frac{BP - 75.49196}{23.55879} \)
Compute \( aGH = \frac{GH - 72.21316}{20.16964} \)
Compute \( aVT = \frac{VT - 61.05453}{20.86942} \)
Compute \( aSF = \frac{SF - 83.59753}{22.37642} \)
Compute \( aRE = \frac{RE - 81.29467}{33.02717} \)
Compute \( aMH = \frac{MH - 74.84212}{18.01189} \)

Compute \( \text{AAGGPHYS} = (aPF \times 0.42402) + (aRP \times 0.35119) + (aBP \times 0.31754) + 
(aGH \times 0.24954) + (aVT \times 0.02877) + (aSF \times -0.00753) + (aRE \times -0.19206) + (aMH \times -0.22069) \)

Compute \( \text{AAGGMENT} = (aPF \times -0.22999) + (aRP \times -0.12329) + (aBP \times -0.09731) + 
(aGH \times -0.01571) + (aVT \times 0.23534) + (aSF \times 0.26876) + (aRE \times 0.43407) + 
(aMH \times 0.48581) \)

Compute \( \text{SF-36 Physical (PCS)} = 50 + (\text{AAGGPHYS} \times 10) \)
Compute \( \text{SF-36 Mental (MCS)} = 50 + (\text{AAGGMENT} \times 10) \)
APPENDIX 6 PSYCHOMTERIC OUTCOME MEASURES

(NB: These questionnaires are not reproduced in exactly the same format as those which were sent to respondents but all relevant questions are included)

How would you rate your facial/dental problems now? (Questionnaires 1, 2 & 3) (mark with a cross on the line e.g.  ________________ x ________________)

10 ________________________________________ 0

<table>
<thead>
<tr>
<th>worst problem</th>
<th>no problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>imaginable</td>
<td>at all</td>
</tr>
</tbody>
</table>

Compared with one year ago, how would you rate your dental/ facial appearance and function now? (Questionnaire 3 only)

1. Much better than 1 year ago
2. Somewhat better than 1 year ago
3. About the same as 1 year ago
4. Somewhat worse than 1 year ago
5. Much worse than 1 year ago

Also included: OQLQ and SF-36
### SPIELBERGER’S STATE AND TRAIT ANXIETY (Questionnaire 1, 2 & 3)

A number of statements which people use to describe themselves are given below. Read each statement carefully and then indicate, using the scale below, how you feel **RIGHT NOW/AT THIS MOMENT**. Circle whichever number best indicates how you feel right now. There are no right or wrong answers. Do not spend too long on any one question. Please answer all questions.

<table>
<thead>
<tr>
<th>NOT AT ALL (1)</th>
<th>SOMEWHAT (2)</th>
<th>MODERATELY (3)</th>
<th>VERY MUCH (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I feel calm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I feel secure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I am tense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I am regretful</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I feel at ease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I feel upset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I am presently worrying over possible misfortunes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I feel rested</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I feel anxious</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I feel comfortable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I feel self-confident</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. I feel nervous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. I am jittery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. I feel “highly strung”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. I am relaxed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. I feel content</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. I am worried</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Read each statement carefully and then indicate, using the scale below, how you have been feeling *generally in the past few weeks*. Circle whichever number best indicates how you feel right now. There are no right or wrong answers.

<table>
<thead>
<tr>
<th>NOT AT ALL</th>
<th>SOMEWHAT</th>
<th>MODERATELY</th>
<th>VERY MUCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

1. I feel pleasant
2. I tire quickly
3. I feel like crying
4. I wish I could be as happy as others seem
5. I am losing out on things because I can't make up my mind soon enough
6. I feel rested
7. I am "cool, calm and collected"
8. I feel that difficulties are piling up so that I can not overcome them
9. I worry too much over something that really doesn't matter
10. I am happy
11. I am inclined to take things hard
12. I lack self-confidence
13. I feel secure
14. I try to avoid facing a crisis or difficulty
15. I feel blue/ down/ fed up
16. I am content 1 2 3 4
17. Some unimportant thing runs through my mind and bothers me 1 2 3 4
18. I take disappointments so keenly that I can't put them out of my mind 1 2 3 4
19. I am a steady person 1 2 3 4
20. I get in a state of tension or turmoil as I think over my recent concerns and interests 1 2 3 4

SPIELBERGER’S STATE AND TRAIT ANXIETY - SCORING SYSTEM

State Anxiety
Reverse Score 1, 2, 5, 8, 10, 11, 15, 16, 19, 20

Trait Anxiety
Reverse Score 1, 6, 7, 10, 13, 16, 19
BECK DEPRESSION INVENTORY (Questionnaire 1, 2 & 3)

This questionnaire consists of 20 groups of statements. After reading each group of statements, circle the number (0, 1, 2, 3) next to the one statement in each group which best describes the way you have been feeling during the past week including today. If several statements in each group seem to apply, circle each one.

*Read all statements in each group before you make your choice. Please answer all questions.*

1.  
   0  I do not feel sad  
   1  I feel sad  
   2  I am sad all the time and I can’t snap out of it  
   3  I am so sad or unhappy that I can’t stand it

2.  
   0  I am not particularly discouraged about the future  
   1  I feel discouraged about the future  
   2  I feel that I have nothing to look forward to  
   3  I feel that the future is hopeless and that things cannot improve

3.  
   0  I do not feel like a failure  
   1  I feel I have failed more than the average person  
   2  As I look back on my life all I can see is a lot of failures  
   3  I feel I am a complete failure as a person

4.  
   0  I get as much satisfaction out of things as I used to  
   1  I don’t enjoy things the way I used to  
   2  I don’t get real satisfaction out of anything anymore  
   3  I am dissatisfied or bored with everything

5.  
   0  I don’t feel particularly guilty  
   1  I feel guilty a good part of the time  
   2  I feel quite guilty most of the time  
   3  I feel guilty all of the time

6.  
   0  I don’t feel I am being punished  
   1  I feel I may be punished  
   2  I expect to be punished  
   3  I feel I am being punished

7.  
   0  I don’t feel disappointed in myself  
   1  I am disappointed in myself  
   2  I am disgusted with myself  
   3  I hate myself

8.  
   0  I don’t feel I am any worse than anyone else  
   1  I am critical of myself for my weaknesses or mistakes  
   2  I blame myself all the time for my faults  
   3  I blame myself for everything bad that happens
<table>
<thead>
<tr>
<th>Question</th>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. I don't cry anymore than usual</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. I cry more now than I used to</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. I cry all the time now</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I used to be able to cry but now I can't cry even though I want to</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>10. I am no more irritated now than I ever am</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. I get annoyed or irritated more easily than I used to</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. I feel irritated all the time now</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I don't get irritated at all by the things that used to irritate me</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>11. I have not lost interest in other people</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. I am less interested in other people than I used to be</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. I have lost most of my interest in other people</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I have lost all of my interest in other people</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>12. I make decisions as well as I ever could</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. I put off making decisions more than I used to</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. I have greater difficulty in making decisions than before</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I can't make decisions at all anymore</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>13. I don't feel I look any worse than I used to</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. I am worried that I am looking old or unattractive</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. I feel that there are permanent changes to my appearance that make me look unattractive</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I believe that I look ugly</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>14. I can work about as well as before</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. It takes an extra effort to get started at doing something</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. I have to push myself very hard to do anything</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I can't do any work at all</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>15. I can sleep as well as usual</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. I don't sleep as well as I used to</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. I wake up 1-2 hours earlier than usual and find it hard to get back to sleep</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I wake up several hours earlier than I used to and cannot get back to sleep</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>16. I don't get more tired than usual</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. I get tired more easily than I used to</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. I get tired from doing almost anything</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I am too tired to do anything</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>17. My appetite is no worse than usual</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1. My appetite is not as good as it used to be</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2. My appetite is much worse now</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3. I have no appetite at all anymore</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
18.  
0  I haven’t lost much weight, if any, lately  
1  I have lost more than 5 pounds (2 kilos)  
2  I have lost more than 10 pounds (4 kilos)  
3  I have lost more than 15 pounds (7 kilos)  

I am purposely trying to lose weight by eating less:— YES _________ NO _________

19.  
0  I am no more worried about my health than usual  
1  I am worried about physical problems such as aches and pains or upset stomach  
2  I am very worried about physical problems and it’s hard to think of anything else  
3  I am so worried about my physical problems that I cannot think of anything else  

20.  
0  I have not noticed any recent changes in my interest in sex  
1  I am less interested in sex than I used to be  
2  I am much less interested in sex than I used to be  
3  I have lost interest in sex completely  

BECK DEPRESSION INVENTORY - SCORING SYSTEM  
Scores for individual items are added
SARASON'S SOCIAL SUPPORT QUESTIONNAIRE (1st Questionnaire only)

The following questions ask about people in your environment who provide you with help or support. Each question has two parts. For the first part, list all the people you know, excluding yourself, whom you can count on for help or support in the manner described. Give each person’s initials and their relationship to you (see example). Do not list more than one person next to each of the numbers beneath each question. Do not list more than nine people per question.

For the second part, using the scale below, circle how satisfied you are with the overall support you have.

6 5 4 3 2 1
Very satisfied Fairly satisfied A little satisfied A little dissatisfied Fairly dissatisfied Very dissatisfied

If you have no support for a question, tick the words “No one”, but still rate your level of satisfaction.

Example

Who do you know whom you can trust with information that could get you in trouble?

(a) No one
   1) SA (mother) 4) RW (friend) 7)
   2) PA (brother) 5) 8)
   3) TL (boss) 6) 9)

(b) How satisfied?
   6 5 4 3 2 1

QUESTIONS :-

1. Whom can you really count on to distract you from your worries when you feel under stress?

(a) No one
   1) 4) 7)
   2) 5) 8)
   3) 6) 9)

(b) How satisfied?
   6 5 4 3 2 1
2. Whom can you really count on to help you feel more relaxed when you are under pressure or tense?

(a) No one
1) 4) 7)
2) 5) 8)
3) 6) 9)

(b) How satisfied?
6 5 4 3 2 1

3. Who accepts you totally, including both your worst and best points?

(a) No one
1) 4) 7)
2) 5) 8)
3) 6) 9)

(b) How satisfied?
6 5 4 3 2 1

4. Whom can you really count on to care about you, regardless of what is happening to you?

(a) No one
1) 4) 7)
2) 5) 8)
3) 6) 9)

(b) How satisfied?
6 5 4 3 2 1

5. Whom can you really count on to help you feel better when you are feeling generally down-in-the-dumps?

(a) No one
1) 4) 7)
2) 5) 8)
3) 6) 9)

(b) How satisfied?
6 5 4 3 2 1
APPENDIX 6  QUESTIONNAIRES

6. Whom can you count on to console you when you are upset?
   (a) No one
   1) 7) 4) 2) 8) 3) 9)
(b) How satisfied?
   6  5  4  3  2  1

SARASON'S SHORT FORM SOCIAL SUPPORT QUESTIONNAIRE SCORING

Number of supports
Add individual scores (0-54) and divide by 6 to give mean score

Satisfaction with supports
Add individual scores (6-36) and divide by 6 to give mean score
SECORD AND JOURARD'S BODY CATHEXIS SCALE (Questionnaire 1, 2 & 3)
These questions ask about the images you have of yourself. Consider each item and place a tick (✓) in the column which best represents how you feel about yourself at the moment.

<table>
<thead>
<tr>
<th></th>
<th>I have strong feelings about this aspect of my appearance - I wish I could change it</th>
<th>I don't like it but I can put up with it</th>
<th>I have no particular feelings one way or the other</th>
<th>I am satisfied with this aspect of appearance</th>
<th>I consider myself very fortunate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Height</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Hair</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Ears</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Forehead</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Eyes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Nose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Lips</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Mouth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Teeth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Facial complexity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Chin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Neck</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Shoulders</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Chest (males)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Breasts (females)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Arms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>Hands</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>Waist</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>Hips</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>Thighs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>Legs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>General muscle tone development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>Overall facial attractiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>Overall upper body attractiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>Overall lower body attractiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>Overall body appearance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Secord and Jourard's Body Cathexis Scale - Scoring System

<table>
<thead>
<tr>
<th>Strong feelings</th>
<th>Consider myself fortunate</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Add individual item scores

### Facial Body Image

Items: 5, 6, 7, 8, 9, 10, 11, 12, 14, 24

Add individual item scores
ROSENBERG’S SELF ESTEEM SCALE (Questionnaire 1, 2 & 3)

Here is a list of statements dealing with your general feelings about yourself.

If you **strongly agree** with the statement circle SA

If you **agree** with the statement circle A

If you **disagree** with the statement circle D

If you **strongly disagree** with the statement circle SD

<table>
<thead>
<tr>
<th></th>
<th>1 Strongly agree</th>
<th>2 Agree</th>
<th>3 Disagree</th>
<th>4 Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>On the whole, I am satisfied with myself</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>2.</td>
<td>At times I think I am no good at all</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>3.</td>
<td>I feel that I have a number of good qualities</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>4.</td>
<td>I am able to do things as well as most other people</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>5.</td>
<td>I feel I do not have much to be proud of</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>6.</td>
<td>I certainly feel useless at times</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>7.</td>
<td>I feel that I am a person of worth, at least on an equal plane with others</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>8.</td>
<td>I wish I could have more respect for myself</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>9.</td>
<td>All in all, I’m inclined to feel that I’m a failure</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>10.</td>
<td>I take a positive attitude to myself</td>
<td>SA</td>
<td>A</td>
<td>D</td>
</tr>
</tbody>
</table>

ROSENBERG’S SELF ESTEEM SCALE - SCORING SYSTEM

**Positive self esteem**

Items 1, 3, 4, 7, 10  
1=1; 2=2; 3=3; 4=4

**Negative self esteem**

Items 2, 5, 6, 8, 9  
1=4; 2=3; 3=2; 4=1
REFERENCES
Albandar J.M. (1990)
Some predictors of radiographic alveolar bone height reduction over 6 years.

Radiographic quantification of alveolar bone level changes. Predictors of longitudinal bone loss.

Multi-level statistical models in studies of periodontal diseases.

Do item weights matter? An assessment using the Oral Health Impact Profile.
Community Dent Health 14: 133-138.

A comparison of the validity of generic- and disease-specific measures in the assessment of oral health-related quality of life.

Quality of life: a dynamic construct.

Social indicators of well-being: Americans’ perceptions of life quality.
New York: Plenum Press.

Healthy decision making: a new approach in health promotion using health state utilities.
Community Dent Health 12: 8-11.

Federal Register 58: 4601-4614.

The General Oral Health Assessment Index.
Chapel Hill: University of North Carolina, Dental Ecology.
A multilevel analysis of factors affecting pocket probing depth in patients responding differently to periodontal treatment.
J Clin Periodontol 26: 67-76.

Inconsistent responses in three preference-elicitation methods for health states.

The initial development of a health-related quality of life model as an outcome measure in epilepsy.

Bakker C. and van der Linden S. (1995)
Health related utility measurement: an introduction.
J Rheumatol 22: 1197-1199.

Psychological evaluation of patients scheduled for orthognathic surgery.

Barrera M. Jr. and Ainlay S.L. (1983)
The structure of social support: a conceptual and empirical analysis.
J Community Psychol 11: 133-143.

An inventory for measuring depression.
Arch Gen Psychiatry 4: 561-571.

Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation.
Clin Psychol Rev 8: 77-100.

Appearance and the influence of reconstructive surgery on body image.

Measurement of patient-derived utility values for periodontal health using a multi-attribute scale.
REFERENCES

The Sickness Impact Profile: development and final revision of a health status measure.  
Med Care 19: 787-805.

Oxford: Oxford University Press.

What do patients value? Willingness to pay for ultrasound in normal pregnancy.  
Med Care 23: 881-893.

The management and utilisation of operating departments.  
London: NHS Management Executive VFM Unit.

Statistical methods for assessing agreement between two methods of clinical measurement.  

Bleichrodt H. and Johannesson M. (1997)  
An experimental test of a theoretical foundation for rating-scale valuations.  

Measuring health-related quality of life outcomes in women with endometriosis - results of the Gynaecology Audit Project in Scotland.  

Measuring disease.  
Buckingham: Open University Press.

Measuring health. (2nd edition)  
Buckingham: Open University Press.

A comparison of methods of assessing voice quality in laryngeal cancer.  
Toronto: Department of Health Administration, University of Toronto.
Economic evaluation of neonatal intensive care of very-low-birth-weight infants.  

Developing multiattribute health indexes.  
Med Care 22: 1045-1057.

Validating the SF-36 health survey questionnaire: new outcome measure for primary care.  

Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire.  

A review of the use of health status measures in economic evaluation.  

The relevance of health state utilities to lower third molar surgery.  

Briggs A. (1999)  
Economic notes: handling uncertainty in economic evaluation.  
Br Med J 319: 120.

The development of an index of orthodontic treatment priority.  

Strategies and pitfalls in quality of life research.  
Hepatology 29 (6 Suppl) 9S-12S.

Quality of life as a construct in health and disability research.  

Disease-specific and generic health outcomes: a model for the evaluation of long-term intrathecal opioid therapy in noncancer low back pain patients.  
Bulpitt C.J. (1997)
Quality of life as an outcome measure.

Modeling longitudinal mandibular growth: percentiles for gnathion from 6 to 15 years of age in girls.

Measurement of quality of life in surgery.

Allocating resources to health care: is the QALY (Quality Adjusted Life Year) a technical solution to a political problem?

Economic evaluation of the cochlear implant.

Chartered Institute of Public Finance and Accountancy - CIPFA (1998)
The Health Service Database 1998.
Croydon: Institute of Public Finance Ltd.

Cost-utility of the cochlear implant in adults: a meta analysis.

Measuring heart patients' willingness to pay for changes in angina symptoms.

Future directions of quality of life research (Chapter 10).
In: Spilker B., Ed. Quality of Life and Pharmacoeconomics in Clinical Trials, 2nd ed.

Cohen S. (1988)
Psychosocial models of the role of social support in the etiology of physical disease.
Health Psychol 7: 269-297.

Assigning values to intermediate health states for cost-utility analysis: theory and practice.
Stress, social support and the buffering hypothesis.  
Psychol Bull 98: 310-357.

Development of a health-related quality-of-life questionnaire for individuals with gastroesophageal reflux disease: a validation study.  
Dig Dis Sci 44: 1376-1383

Oxford: Oxford University Press

Coopersmith S. (1981)  
The antecedents of self-esteem.  

Measuring oral health: does your treatment really make a difference?  

Development of a health-related quality-of-life questionnaire (PCOSQ) for women with polycystic ovary syndrome (PCOS).  

Survey of patients’ attitudes after surgical correction of prognathism: study of 33 patients.  

Perceptions of outcome following orthognathic surgery.  

Cunningham S.J., Gilthorpe M.S. and Hunt N.P. (2000)  
Are orthognathic patients different?  

Oxford: Oxford University Press.

Developing socio-dental indicators - the social impact of dental disease.  
Community Dent Health 3: 3-17.
A prospective study on quality of life of laryngeal cancer patients treated with radiotherapy.

de Lateur B.J. (1997)

Department of Health (1994)
An oral health strategy for England.
London: Department of Health.

Department of Health (1995)
Policy appraisal and health.
London: Department of Health.

Surveys in social research. (4th ed)

Dijkers M. (1999)
Measuring quality of life: methodological issues.

Dolan P. (1999)
Valuing health-related quality of life. Issues and controversies.
Pharmacoeconomics 15: 119-127.

The time trade-off method: results from a general population study.
Health Econ 5: 141-154.

Donaldson C. (1990a)
The state of the art of costing health care for economic evaluation.
Community Health Stud 14: 341-356.

Donaldson C. (1990b)
Willingness to pay for publicly-provided goods. A possible measure of benefit?
J Health Econ 9: 103-118.

Donaldson C. (1993)
Theory and practice of willingness to pay for health care.
Health Economics Research Unit, University of Aberdeen, Discussion Paper 01/93.
REFERENCES

Donaldson C., Hundley V. and Mapp T. (1995a)
Willingness to pay: a new method for measuring patients’ preferences?
Health Economics Research Unit, University of Aberdeen, Discussion Paper 03/95.

Donaldson C., Thomas R. and Torgerson D.J. (1995b)
Open-ended versus payment scale approaches to eliciting willingness to pay.
Health Economics Research Unit, University of Aberdeen, Discussion Paper 01/95.

Estimating the economic benefits of avoiding food-borne risk: is willingness to pay feasible?

Donaldson C. and Shackley P. (1997)
Does “process utility” exist? A case study of willingness to pay for laparoscopic cholecystectomy.

Downer M.C., Jullien J.A. and Speight P.M. (1997)
An interim determination of health gain from oral cancer and precancer screening: 1.
Obtaining health state utilities.
Community Dental Health 14: 139-142.

Health gain from restorative dental treatment evaluated by computer simulation.

Methods for the economic evaluation of health care programmes.
Oxford: Oxford University Press.

Drummond M.F. and Davies L. (1991)
Economic analysis alongside clinical trials. Revisiting the methodological issues.

Cost-effectiveness league tables: think of the fans.
Health Policy 31: 231-238.

Evaluation of the dyspeptic patient: a cost-utility study.
J Fam Pract 44: 545-555.

Edgerton M.T. Jr. and Knorr N.J. (1971)
Motivational patterns of patients seeking cosmetic (esthetic) surgery.
The cost, effectiveness and cost effectiveness of removal and retention of
asymptomatic, disease free third molars.

An introduction to the bootstrap.
New York: Chapman and Hall.

Epker B.N. and Wolford L.M. (1975)
Middle-third facial osteotomies: their use in the correction of acquired and
developmental dentofacial and craniofacial deformities.

EuroQol Group (1990)
EuroQol - a new facility for the measurement of health-related quality of life. The
EuroQol Group.
Health Policy 16: 199-208.

Surgical day care: measurements of the economic payoff.

Preliminary evaluation of an illustrated scale for rating dental attractiveness.

Incorporating utility-based quality-of-life assessment measures in clinical trials. Two
examples.
Med Care 27 (3 suppl): S190-204.

Outcome assessment of implant-supported prostheses.

Feinmann C. (1999)
Personal communication

An introduction to utility measurement in health care.

Ferrans C.E. and Powers M.J. (1985)
Quality of life index: development and psychometric properties.
REFERENCES

Orthognathic surgery: patient expectations; psychological profile and satisfaction with outcome.

Quality of life measures in health care. I: Applications and issues in assessment.

Flanary C.M., Barnwell G.M. Jr. and Alexander J.M. (1985)
Patient perceptions of orthognathic surgery.

Impact of orthognathic surgery on normal and abnormal personality dimensions: a 2 year follow-up study of 61 patients.

Population based cost utility study of interferon beta-1b in secondary progressive multiple sclerosis.

Froberg D.G. and Kane R.L. (1989a)

Froberg D.G. and Kane R.L. (1989b)
Methodology for measuring health-state preferences - III: Population and context effects.

Furlong W., Feeny D., Torrance G.W., Barr R. and Horsman J. (1990)
Guide to design and development of health-state utility instrumentation.

Fyffe H.E. and Kay E.J. (1992)
Assessment of dental health state utilities.

Willingness-to-pay as a measure of benefits. Relevant questions in the context of public decision making about health care programs.
Med Care 29: 1246-1252.

Gafni A. (1994)
The standard gamble method: what is being measured and how it is interpreted.
Health Serv Res 29: 207-224.

Statistics with confidence.

Garratt A.M. (1997)
Personal communication

The SF 36 health survey questionnaire: an outcome measure suitable for routine use within the NHS?

SF 36 health survey questionnaire: II. responsiveness to changes in health status in four common clinical conditions.
Qual Health Care 3: 186-192.

Responsiveness of the SF-36 and a condition-specific measure of health for patients with varicose veins.

Developing a condition-specific measure of health for patients with dyspepsia and ulcer-related symptoms.

QALY league tables: handle with care.
Health Econ 2: 59-64.

Gerard K., Dobson M. and Hall J. (1993)
Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer.
J Clin Epidemiol 46: 77-84.
Assessing the performance of utility techniques in the absence of a gold standard.
Med Care 37: 580-588.

Gift H.C. and Redford M. (1992)
Oral health and the quality of life.

Oral health, health and health-related quality of life.
Med Care 33 (11 suppl): NS57-77.


Gilthorpe M.S. and Cunningham S.J. (2000)
The application of multilevel modelling to multivariate dental research data.
Community Dent Health.
In press.

Goldberg D. and Williams P. (1988)
A user's guide to the General Health Questionnaire.
Windsor, Berkshire: NFER-Nelson Publishing Company Ltd.

Goldstein H. (1987)
Multilevel models in educational and social research.
London: Oxford University Press.

Multilevel statistical models.
London: Edward Arnold.

A user's guide to MLwiN.
Multilevel Models Project, Institute of Education, University of London.

Allocating health care: cost-utility analysis, informed democratic decision making, or the veil of ignorance?
Patient utilities in chronic musculoskeletal pain: how useful is the standard gamble method?
Pain 80: 365-375.

Cost effectiveness of anterior cruciate ligament reconstruction in young adults.

The QALY toolkit. Discussion Paper 38
Centre for Health Economics, Health Economics Consortium, University of York.

Health state valuations from the general public using the visual analogue scale.
Qual Life Res 5: 521-531.

Guyatt G. (1997)
Insights and limitations from health-related quality-of-life research.

A measure of quality of life for clinical trials in chronic lung disease.

A new measure of health status for clinical trials in inflammatory bowel disease.
Gastroenterology 96: 804-810.

Measuring health-related quality of life.

Generic and specific measurement of health-related quality of life in a clinical trial of respiratory rehabilitation.

Hadorn, D.C. (1991)
Setting health care priorities in Oregon. Cost-effectiveness meets the rule of rescue.
Hall J., Gerard K., Salkeld G. and Richardson J. (1992)
A cost-utility analysis of mammography screening in Australia.

Components and predictors of patient satisfaction.
Br J Health Psychol 1: 65-85

Hatch J.P., Rugh J.D., Clark G.M., Keeling S.D., Tiner B.D. and Bays R.A.
(1998)
Health-related quality of life following orthognathic surgery.

Psychol Med 10: 723-734.

Choice of long-term strategy for the management of patients with severe esophagitis:
a cost-utility analysis
Gastroenterology 112: 1078-1086.

Hine D. (1999)
For the good that it will do: issues confronting healthcare in the UK.

Hoeksma J.B. and van der Beek M.C.J. (1991)
Multilevel modelling of longitudinal cephalometric data explained for orthodontists.

Holman A.R., Brumer S., Ware W.H. and Pasta D.J. (1995)
The impact of interpersonal support on patient satisfaction with orthognathic surgery.

Hornberger J.C., Redelmeier D.A. and Petersen J. (1992)
Variability among methods to assess patients’ well-being and consequent effect on a
cost-effectiveness analysis.

Med Care 34 (12 suppl.): DS23-DS33.

Hunt S.M. and McKenna S.P. (1992)
The QLDS: a scale for the measurement of quality of life in depression.
Hunt N.P. and Cunningham S.J. (1997)
The influence of orthognathic surgery on occlusal force in patients with vertical facial deformities.

Hutton C.E. (1967)
Patients’ evaluation of surgical correction of prognathism: survey of 32 patients.

A scale for assessing quality of life in adult asthma sufferers.

Domains, constructs and the development of the breathing problems questionnaire.

A utility analysis of dental implants.

James M., St Leger S. and Rowsell K.V. (1996)
J Epidemiol Community Health 50: 182-189.

Quality of Life Measurement in Health Care: A review of measures and population norms for the UK SF-36.
University of Oxford Press

Health-related quality-of-life measurement in patients with Parkinson’s disease.
Pharmacoeconomics 15: 157-165.

Changing values in medical and health care decision making.
John Wiley and Sons.

Johannesson M. (1994)
The concept of cost in the economic evaluation of health care. A theoretical inquiry.

Development and testing of a new measure of health status for clinical trials in rhinoconjunctivitis.
Clin Exp Allergy 21: 77-83.
How to develop and validate a new health-related quality of life instrument (Chapter 6).
In: Spilker B., Ed. Quality of life and pharmacoeconomics in clinical trials, 2nd ed.

Clinical impact versus factor analysis for quality of life questionnaire construction.
J Clin Epidemiol 50: 233-238.

Development and validation of the Mini Asthma Quality of Life Questionnaire.

Kahneman D. and Tversky A. (1979)
Prospect Theory: an analysis of decision under risk.

Oxford: Oxford University Press.

Self-derogation and psychosocial adjustment.

Valuation of health changes with the contingent valuation method: a test of scope and question order effects.
Health Econ 5: 531-541.

Kind P. (1996)
The EuroQol Instrument: an index of health-related quality of life (Chapter 22).
In: Spilker B., Ed. Quality of life and pharmacoeconomics in clinical trials, 2nd ed.

Personal communication

Sex differences in motives for and outcomes of orthognathic surgery.
Predicting psychological responses to orthognathic surgery.

The psychological impact of orthognathic surgery: A 9-month follow-up.

Psychological changes in orthognathic surgery patients: A 24-month follow-up.

The emotional impact of orthognathic surgery and conventional orthodontics.
Am J Orthod 88: 224-234

Personality characteristics as predictors and sequelae of surgical and conventional orthodontics.

Klock B. (1980)
Economic aspects of a caries preventive program.

Konstantinovic V.S. (1999)
Quality of life after surgical excision followed by radiotherapy for cancer of the tongue and floor of the mouth: evaluation of 78 patients.

Krischer J.P. (1976)
Utility structure of a medical decision-making problem.

Theoretical and empirical considerations on the relation between “body image”, body scheme and somatosensation.

Lawton M.P. (1999)
Quality of life in chronic illness.
Gerontology 45: 181-183.
What is the cost utility of screening for hepatitis C virus (HCV) in intravenous drug users?

Liang MH, Fossel AH and Larson MG. (1990)
Comparisons of five health status instruments for orthopedic evaluation.
Med Care 28: 632-642.

Locker D. (1997)
Concepts of oral health, disease and the quality of life.
Chapel Hill: University of North Carolina, Dental Ecology.

The specific psychosocial effects of orthognathic surgery.

The economics of complete dentures and implant-related services: a framework for analysis and preliminary outcomes.

Mason J. and Drummond M. (1995)
Discussion Paper 128, Centre for Health Economics, University of York.

Assessing health-related quality-of-life and health state preference in persons with obesity: a validation study.
Qual Life Res 6: 311-322

Preference based measurements in dentistry: a review of the literature and recommendations for research.
Community Dent Health 16: 5-11.

Measuring health - a guide to rating scales and questionnaires.
Oxford: Oxford University Press.
The Nottingham Health Profile (Chapter 34).
In: Spilker B., Ed. Quality of life and pharmacoconomics in clinical trials, 2nd ed.

McGrath C. and Bedi R. (1999)
The importance of oral health to older people's quality of life.
Gerodontontology 16: 59-63

The economics of health care: an introductory text.

Providing health care: The economics of alternative systems of finance and delivery.
Oxford: Oxford University Press.

The MOS 36-Item Short-Form Health Survey (SF-36): II. psychometric and clinical
tests of validity in measuring physical and mental health constructs.
Med Care 31: 247-263.

Quality-adjusted life years, utility theory, and healthy-years equivalents.
Med Decis Making 9: 142-149.

The healthy-years equivalents: how to measure them using the standard gamble
approach.
Med Decis Making 11: 140-146.

evaluation and selection of standardized psychological instruments for research and
clinical practice.

Miller A.J. (1990)
Subset selection in regression.
Chapman and Hall, London.

Mishan E.J. (1988)
London: Unwin Hyman.

Using surveys to value public goods: the contingent valuation method.
Utility scores for chronic conditions in a community-dwelling population.
Pharmacoeconomics 15: 369-376.

Mjör IA. (1992)
Long term cost of restorative therapy using different materials.

The relative cost of different restorations in the UK.

Mooney G. (1994)
Key issues in health economics.
Harvester Wheatsheaf.

Economic evaluation of a pit and fissure dental sealant and fluoride mouthrinising program in two nonfluoridated regions of Victoria, Australia.

Morrison G.C. and Gyldmark M. (1992)
Appraising the use of contingent valuation.
Health Econ 1: 233-243.

Muldoon M.F., Barger S.D., Flory J.D., Manuck S.B. (1998)
What are quality of life measurements measuring?

A comparison of body image perceptions in bulimics, restrainers and normal women: An extension of previous findings.
Addict Behav 18: 503-509.

NHS Centre for Reviews and Dissemination, University of York (1999)
Dental restoration: what type of filling?
Effective Health Care 5: 1-12.

NHS Executive (1999)
Health Service Cost Index: HSHS specific price inflation.
Leeds: NHSE.

Incorporating societal concerns for fairness in numerical valuations of health programmes.
Health Econ 8: 25-39.
Psychometric theory.  

Individual quality of life in patients undergoing hip replacement.  

Willingness-to-pay: A valid and reliable measure of health state preference?  
Ontario, Canada: Center for Health Economics and Policy Analysis (CHEPA)  
Working Paper Series # 93-7, McMaster University.

Willingness to pay: a valid and reliable measure of health state preference?  

When do the “dollars” make sense? Toward a conceptual framework for contingent valuation studies in health care.  

Assessing oral health outcomes for orthodontics - measuring health status and quality of life.  

Helicopters, hearts and hips: using willingness to pay to set priorities for public sector health programmes.  

O'Young J. and McPeek B. (1987)  
Quality of life variables in surgical trials.  
J Chronic Dis 40: 513-522.

A prospective study of the cost-utility of the multichannel cochlear implant.  

Patient satisfaction in primary health care: a literature review and analysis.  

Generic and disease-specific measures in assessing health status and quality of life.  
Med Care 27: (3 Suppl) S217-232.
Psychological considerations in corrective maxillary and midfacial surgery.

The development and validation of a short measure of functioning and well-being for individuals with Parkinson's disease.

Statistics for veterinary and animal science.
London: Blackwell Science Ltd.

The QALY: A guide for the public health physician.

Pinto Prades J-L. (1997)
Is the person trade-off a valid method for allocating health care resources?
Health Econ 6: 71-81.

Surgery versus stereotactic radiosurgery for small, operable cerebral arteriovenous malformations: a clinical and cost comparison.
Neurosurgery 41: 757-764.

The development of a measure of social support: The Significant Others (SOS) Scale.

The effect of orthognathic surgery on occlusal force.

Orthognathic surgery: a hierarchy of stability.
Int J Adult Orthodont Orthognath Surg 11: 191-204

Rachel (1999)
Rachel's story.
Br J Orthod 26: 244-246.

A user's guide to MLwiN.
Institute of Education, London.
Read J.L. (1993)  
The new era of quality of life assessment (Chapter 1).  

Impact of dental conditions on patients' quality of life.  
Community Dent Oral Epidemiol 17: 7-10.  

Oxford: Oxford University Press.  

Rice J.A. (1988)  
Mathematical statistics and data analysis.  

Richardson J. (1994)  
Cost utility analysis: what should be measured?  

The development of the PAR Index (Peer Assessment Rating): reliability and validity.  
Eur J Orthod 14: 125-139.  

Int J Technol Assess Health Care 11: 365-376  

Robinson R. (1993a)  
Economic analysis and health care. What does it mean?  

Robinson R. (1993b)  
Costs and cost-minimisation analysis.  

Robinson R. (1993c)  
Cost-effectiveness analysis.  
REFERENCES

Robinson R. (1993d)
Cost-utility analysis.

Robinson R. (1993e)
Cost-benefit analysis.

Rogers S.N., Fisher S.E. and Woolgar J.A. (1999a)
A review of quality of life assessment in oral cancer.

Quality of life 5-10 years after primary surgery for oral and oro-pharyngeal cancer.

Developing measures of patients' perceptions of orthognathic surgery.

Rosenberg M. (1965)
Society and the adolescent self-image.

A new approach to the measurement of quality of life. The Patient-Generated Index.
Med Care 32: 1109-1126.

The utility of different health states as perceived by the general public.

Assessing social support: the social support questionnaire.

A brief measure of social support. Practical and theoretical implications.

Quality of life studies: definitions and conceptual issues (Chapter 2).
In: Spilker B., Ed. Quality of life and pharmacoeconomics in clinical trials, 2nd ed.
A comparison of two rating scales for depression.  

A cost-utility analysis of abdominal hysterectomy versus transcervical endometrial resection for the surgical treatment of menorrhagia.  

Personal communication.

An economic evaluation of transcervical endometrial resection versus abdominal hysterectomy for the treatment of menorrhagia.  
Br J Obstet Gynaecol 100: 244-252.

Randomised trial comparing hysterectomy and transcervical endometrial resection: effect on health related quality of life and costs two years after surgery.  
Br J Obstet Gynaecol 103: 142-149.

Secord P.F. and Jourard S.M. (1953)  
The appraisal of body cathexis: body cathexis and the self.  
J Consult Psychol 17: 343-347.

Health state utilities in dentistry: A review.  

Short-term cost-effectiveness analysis of presurgical orthopedic treatment in children with complete unilateral cleft lip and palate.  
Cleft Palate Craniofac J 35: 222-226.

Derivation and validation of a short-form Oral Health Impact Profile.  

Assessing change in quality of life using the Oral Health Impact Profile.  
Community Dent Oral Epidemiol 26: 52-61.

Development and evaluation of the Oral Health Impact Profile.  
Community Dent Health 11: 3-11.
Long-term survival and cost-effectiveness of five dental restorative materials used in various classes of cavity preparations.

A multilevel analysis of factors affecting the longevity of fixed partial dentures, retainers and abutments.

Manual for the State-Trait Anxiety Inventory (revised edition).

Spilker B. (1996)
Introduction (Chapter 1).
In: Spilker B., Ed. Quality of life and pharmaco economics in clinical trials, 2nd ed.

Measuring the quality of life of cancer patients: a concise QL-index for use by physicians.
J Chronic Dis 34: 585-597.

Integrating response shift into health-related quality of life research: a theoretical model.

An empirical note on willingness to pay and starting-point bias.

Variance components analysis of data from periodontal research.

The “utility” of the Time Trade-Off method in cancer patients: feasibility and proportional Trade-Off.
J Clin Epidemiol 48: 1207-1214.
St Leger A.S., Schnieden H. and Walsworth-Bell, J.P. (1991)
Evaluating health services' effectiveness - a guide for health professionals, service
managers and policy makers.
Open University Press.

The Dental Impact Profile.
Chapel Hill: University of North Carolina, Dental Ecology.

Health measurement scales. A practical guide to their development and use. (2nd ed)
Oxford: Oxford University Press.

Thompson M.S. (1986)
Willingness to pay and accept risks to cure chronic disease.

Thompson S.G. and Barber J. (2000)
How should cost data in pragmatic randomised trials be analysed?

Applicability of SWOT analysis for measuring quality of public oral health services as
perceived by adult patients in Finland. Strengths, weaknesses, opportunities and
threats.

Torgerson D.J. and Raftery J. (1999)
Economics notes. Discounting.

Torrance G.W. (1976)
Social preferences for health states: an empirical evaluation of three measurement
techniques.

Preferences for health states: a review of measurement methods.

Torrance G.W. (1986)
Measurement of health state utilities for economic appraisal: a review.
Torrance G.W. (1987)
Utility approach to measuring health-related quality of life.
J Chronic Dis 40: 593-603.

Personal communication.

Torrance G.W., Thomas W.H. and Sackett D.L. (1972)
A utility maximization model for evaluation of health care programs.
Health Serv Res 7: 118-133.

Application of multi-attribute utility theory to measure social preferences for health states.
Oper Res 30: 1043-1069.

Utilities and quality-adjusted life years.
Int J Technol Assess Health Care 5: 559-575.

Townsend P., Phillimore P. and Beattie A. (1986)
Health and deprivation: inequality in the North.
London: Routledge.

Can severity of symptoms be used as an outcome measure in trials of non-ulcer dyspepsia and Helicobacter pylori associated gastritis?

von Neumann J. and Morgenstern O. (1953)
Theory of games and economic behavior. (3rd ed)
New York: Wiley.

Ware J.E. (1993)
Measuring patients’ views: the optimum outcome measure.

Ware J.E. Jr. and Sherbourne C.D. (1992)
Med Care 30: 473-483.

Alternative to mental hospital treatment. II. Economic benefit-cost analysis.
Arch Gen Psychiatry 37: 400-405.
Functioning and utility for current health of patients with depression or chronic medical conditions in managed, primary care practices.
Arch Gen Psychiatry 56: 897-904.

WHOQOL Group (1996)
What quality of life? The WHOQOL Group.
World Health Forum 17: 354-356.

Economics, QALYs and medical ethics - a health economist's perspective.
Health Care Anal 3: 221-226.

Williams L.S., Weinberger M., Harris L.E., Clark D.O. and Biller J. (1999)
Stroke 30: 1362-1369.

Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes.

Wolford L.M. and Epker B.N. (1975)
The combined anterior and posterior maxillary ostectomy: a new technique.

Wong E., Guyatt G.H., Cook D.J., Griffith L.E. and Irvine E.J. (1998)
Development of a questionnaire to measure quality of life in patients with irritable bowel syndrome.

Wood-Dauphinee S. (1999)
Assessing quality of life in clinical research: from where have we come and where are we going?

Younossi Z.M., Guyatt G., Kiwi M., Boparai N. and King D. (1999)
Development of a disease specific questionnaire to measure health related quality of life in patients with chronic liver disease.
Gut 45: 295-300.

Cost-benefit analyses in the health-care literature: don’t judge a study by its label.
The hospital anxiety and depression scale.

Zung W.W.K., Richards C.B. and Short M.J. (1965)
Self-rating depression scale in an outpatient clinic. Further validation of the SDS.
Arch Gen Psychiatry 13: 508-515.
Economic evaluation of healthcare — is it important to us?

S. J. Cunningham, 1

Economic evaluation is now an accepted method for the appraisal of healthcare programmes. Although it is used widely in medicine, its use in the field of dentistry is only just beginning to achieve popularity. Economic evaluation in dentistry is likely to become increasingly important in the future and this paper aims to introduce the basics of the technique as well as describing some of the dental settings in which it is currently being used.

"We never will have all we need. Expectation will always exceed capacity... This service must always be changing, growing and improving, it must always appear inadequate."

Aneurin Bevin, 1948

What is economic evaluation?

The economic evaluation of healthcare programmes has become more important in recent years and this is reflected by an increase in the literature. It is now an accepted tool for the appraisal of healthcare programmes. Studies may be conducted from the viewpoint of individual recipients of healthcare, healthcare providers or society generally1 and such investigations are now being undertaken by researchers from many different fields including economists, medical researchers and clinicians. Economic evaluation may be defined as 'the comparative analysis of alternative courses of action in terms of both their costs and consequences'.2 It involves two main areas, first, the costs and consequences of programmes and, second, choices which have to be made in allocation of resources. Although sometimes viewed with suspicion by both clinicians and the general public, economic evaluation does aim to determine how resources can give the greatest benefit.

Any economic analysis involves measurement of both the benefits of healthcare and also the costs.3 It aims to answer two main questions:

- Is the health procedure in question worth doing compared with other things we could do with the same resources?
- Are we satisfied that the healthcare resources should be spent in this way rather than in any other way?

Benefits may be divided into gains in health status (direct benefits) as well as other indirect benefits (eg production gains). Costs may be divided into direct medical costs (eg costs to the NHS), direct non-medical costs (eg family expenditure, social services) and indirect costs or productivity costs (eg changes associated with treatment such as time off work, earlier return to work). One term which is particularly important is 'opportunity cost', or the value of a resource in its best alternative use, and evaluations therefore aim to compare opportunity costs with the improvement in health as a result of the intervention under examination.4

It is important that those individuals who are involved in the provision and purchase of healthcare fully understand the background to this concept and understand some of the terms which are commonly used. Without good economic analysis, healthcare is unlikely to progress and only by undertaking systematic reviews is it possible to identify alternatives to existing or new programmes. Such evaluation is dependent on the quality of underlying medical evidence and, because of this, clinical trials are now viewed as a natural vehicle for economic analysis.5 However, economic evaluation in healthcare is most useful when certain other questions have already been answered and these include:

1. Can the health procedure/intervention work? (the efficacy of the procedure)
2. Does the procedure/intervention work? (evaluation of effectiveness)
3. Is it reaching those who need it? (availability of the service)

The economic evaluation of healthcare shows certain important differences when compared with other commodities.6 Economics is concerned with choice and the assumption is made that choices are made after the consumer has been provided with full information. However, individuals who present within the healthcare sector are frequently unwilling or unable to collect the information required to make this choice. This is further complicated by the fact that the individual supplying the information is also the person supplying the treatment, something which does not generally happen in other fields. Other areas which make healthcare different include the fact that it is consumed on the day where dentistry may differ from the rest of the healthcare field.

In brief

- Economic evaluation of healthcare programmes is now common-place in medicine and is becoming increasingly important in dentistry.
- Increased research in the field of economic evaluation in conjunction with clinical trials is required in dentistry in both primary care and hospital settings.
- This paper reviews the various methods of economic evaluation and also gives examples of some of the current research in the field of dentistry.
Allocation of resources
The allocation of healthcare resources has always been a controversial issue. Since the introduction of the National Health Service, there has been a basic problem that resources are limited and the demands can not be met. The questions which should always be asked are — is the allocation of healthcare resources efficient? and, is the allocation equitable?

Healthcare and government agencies must decide how to allocate their resources for a wide range of very different health care interventions. This involves making difficult judgments regarding the importance of certain health states. A number of arguments have been proposed in terms of 'need' for healthcare and/or 'right' to treatment. Goold proposed that the only fair way of resource allocation is to have two levels of organisation.7 The first, based on the existing government structure, should be responsible for making broad decisions, such as the amount of money allocated to health care. The second level, made up of community organisations with a membership which represents a wide range of health related interests, would be involved in the making of policy decisions and the development of guidelines.

1. Healthcare is limited by the total amount of resources available as well as through competition with other areas, such as housing and education and it is difficult to determine who should be responsible for the 'rationing' of healthcare. When rationing of resources becomes necessary, some procedure has to be set up to allow the most appropriate allocation. This was the basis for the introduction of cost-utility analysis (see later). There are three main theories which have been proposed to assist the allocation of resources, none of which is without faults.8 Although these theories are unlikely to be useful on an everyday basis, they may aid the development of guidelines:

1. The Utilitarian Theory which argues that healthcare should be distributed so as to maximise the health of society (eg increase life expectancy; reduce infant mortality) without regard to how that good is actually distributed.

2. The Egalitarian Theory which is based on the concept that everyone has a claim to the amount of healthcare resources which gives them a level of health equal to that of others.

3. The Rawlsian Theory9 which proposes that each person has an equal right to the system and when making social and economic choices, those who are least advantaged should have maximum benefit.

Methods of economic evaluation
Robinson,10-15 Drummond et al.,2 and Donaldson9 discussed a number of methods of economic evaluation currently in use: cost-minimisation analysis (CMA), cost-effectiveness analysis (CEA), cost-benefit analysis (CBA) and cost-utility analysis (CUA). These are terms which should be understood as they are likely to be seen more frequently in the literature in the future. It is, however, important to realise that these are only an adjunct to decision making.

Cost-minimisation analysis
This form of analysis is used when the outcomes of two procedures being compared are identical and it is important that the outcomes of the alternative programmes are proven to be the same if the method is used. The aim is usually to find the lowest cost programme and the unit of measurement is cost per intervention. There are few examples in dentistry which would fall into this type of analysis, except perhaps a situation where third molars could be removed as a day case or in-patient stay, with identical outcome. It would then be possible to calculate the costing for both methods and to select the lowest of the two.

Cost-effectiveness analysis
Cost-effectiveness was the most widely used method of economic analysis until the 1980s. It answers the question 'Given that it has been decided that this type of health care will be provided, what is the best way of doing so?' This method is used when the programmes may have differential success in outcome, as well as differential costs, but the outcome must be common to both programmes (eg life years gained; blood pressure reduction). For example, a comparison of several different materials for dental restorations was published recently by the NHS Centre for Reviews and Dissemination.16 In any study such as this, it is not the cheapest programme per se which is of interest, the aim is to find the most efficient treatment option in terms of cost per unit effect (eg cost per tooth year gained).

The disadvantage of the cost-effectiveness approach is that it cannot be used to assess a single programme or to compare interventions which have several different clinical effects. It was this disadvantage which lead to the development of cost-utility analysis (CUA). There are a number of similarities between CEA and CUA and the two terms are sometimes used synonymously.

Cost-benefit analysis
If the outcome of two health programmes differs, then a common denominator must be established to allow comparisons of outcome. One way of doing this is in monetary terms and cost-benefit analysis aims to measure the costs and the consequences in terms of pounds, dollars etc. CBA differs from cost-effectiveness and cost-utility analyses in that costs and benefits of healthcare are expressed in the same units. It is a difficult concept because placing a monetary value on life or relief of suffering is a concept which many find difficult. When assessing a programme in this way, there is usually an assumption that the alternative is to do nothing and this itself may have costs associated with it. It is worth noting that many studies which call themselves 'cost-benefit analyses' are not true CBAs and, in fact, compare the costs of treatment with cost savings.

CBA has a solid foundation in welfare economics and it is one of the most comprehensive methods of economic evaluation which is available. Two main approaches are used. The first is the human capital approach, which values health improvement on the basis of the individual's future 'worth' to society in terms of future earnings. The second method is the so-called 'willingness to pay' approach where respondents are asked how much they would be prepared to pay for a certain form of health
intervention (or alternatively, how much they would be 'willing to accept' to forgo this intervention). It is a technique which has been used widely in environmental issues but has only been used relatively recently in healthcare. Both methods have problems associated with them, particularly the use of the willingness-to-pay method in a society which is not used to paying for healthcare. However, this method may prove useful in analysis of certain forms of dental intervention. For example, comparing the willingness to pay for implant retained prostheses against conventional dentures.

Cost-utility analysis
Utility refers to the value or worth of a particular health state or an improvement in that health state. Utility values lie between 0 and 1, where 0 is equivalent to death and 1 is equivalent to perfect health. CUA should be the method of choice when quality of life is an important outcome. It is also the ideal method when interventions affect both morbidity and mortality or when treatments have a wide range of different outcomes and a common unit is required. Utility values may be estimated using values quoted in the literature or they may be measured directly using a number of techniques such as the Standard Gamble or the Time Trade-Off. In CUA only final data (e.g. lives saved; days of illness avoided) can be used. Intermediate data, such as cases found, cannot be used as they cannot be converted into QALYs gained.

Utility based measures are usually expressed in terms of quality adjusted life years (QALYs) gained, which are calculated by multiplying the change in utility value as a result of medical intervention by the years of life remaining. The cost per QALY is then calculated and this can be used to produce 'league tables' which list interventions in order of cost per QALY. This in turn may be used to guide resource allocation although there is still a great deal of controversy associated with their use. For example, the well known Oregon Study, which aimed to develop a priority list for proposed rationing for individuals on Medicaid, found that the use of splints for temporomandibular joint dysfunction came out higher than appendicectomy. The major surprise here being that treatment for TMD was not only ranked higher than a life saving procedure but that it was also a relatively cheap life saving procedure. It therefore becomes obvious that the information should be handled carefully before making decisions.

The principle behind CUA is that a QALY gained is considered to be worth the same no matter who receives it. Recent papers have suggested that QALYs could be valued differently depending on how seriously ill the individual is. This approach has important implications in the field of dentistry as patients are not seriously ill and treatment may be considered less worthwhile under this system of equity weights.

CUA may be seen as an improvement on CBA as it attempts to combine more than one outcome measure. It may also be seen as an improvement on CBA as it permits ranking of programmes without the need to place monetary values on the benefits. This is a useful method of economic analysis when looking at dental interventions which produce changes in quality of life, for example, improvements following orthodontic treatment or following the placement of implant retained prostheses rather than conventional dentures.

Economic evaluation in dentistry
'When alternative therapies are available, patients want the choice of treatment to be based on processes that are cost-effective and have proven outcomes.' Cost effectiveness and cost benefit analyses of dental interventions by the public and by those funding healthcare and the National Institute for Clinical Excellence (NICE) may play an important role in this area in the future. Both the NHS and private companies are likely to demand increased evidence of value for money in the future. This is particularly important in areas which may be perceived as 'cosmetic'. Economic evaluation is still used less frequently in dentistry than in medicine. However, this is beginning to change. A recent computerised literature search showed the following:

- 'Cost benefit and dentistry' produced a total of 370 papers published between 1971 and 1999 with 66 of those in 1997/98
- 'Cost utility and dentistry' produced only 18 papers, all of which were published between 1980 and 1998 with 5 of those in 1997/98
- 'Cost minimisation and dentistry' produced no papers at all.

It is, however, worth noting that a number of papers were listed under both cost benefit and cost effectiveness. This stresses the importance that papers must be read carefully to determine which method of analysis was actually used. It is also worth noting that a large number of the papers which were listed, had not undertaken any form of economic analysis and merely mentioned that economic evaluation would be a useful next step in research. A relatively small number of the papers had undertaken carefully controlled economic evaluation.

Cost effectiveness and cost benefit studies are therefore carried out much more frequently than cost utility studies, which probably reflects the increased difficulty and time consuming nature of cost utility type studies. However, the cost utility method would be particularly useful in the field of dentistry because treatments frequently produce improvements in quality of life. In addition, QALY based investigations in dentistry would also allow some method of comparing dental interventions with other forms of medicine.

Cost effectiveness and cost benefit studies have focused largely on comparison of restorative materials and cost implications of fluoride, fissure sealants and caries prevention. Recent years have also seen a number of papers undertaking economic analysis of implants. The following examples of economic evaluation in dentistry have been selected to illustrate the issues described in the previous sections. It is not intended to be an exhaustive list.
A good example of clinical trials and economic evaluation being undertaken concurrently is that by Severens et al. who assessed the short-term cost effectiveness of pre-surgical orthopaedics in babies with a complete unilateral cleft of the lip and palate.38 There was a significant difference in both medical and indirect costs for the two groups with the pre-surgical orthopaedic group being higher. However, there was no significant difference in outcome (which was assessed in terms of operating time) between the two groups. Thus concluding that pre-surgical orthopaedics was not cost-effective in terms of reduced operating time. Obviously, other important outcome measures such as appearance and function must be assessed but these were not reported in this paper.

Klock looked at CBA and CEA of a preventive programme (including oral hygiene, fluoride application and fissure sealants) and found that in spite of a reduction in caries activity, the programme was uneconomic compared with traditional dental care.31 In contrast, Morgan et al.33 assessed the cost-effectiveness of a preventive programme in two non-fluoridated regions of Australia and concluded that the introduction of a preventive programme was an efficient use of resources. They also stressed the need for systematic evaluation of a full range of dental prevention and treatment programmes.

A number of cost-effectiveness studies in dentistry have looked at different restorative materials. Mjör studied the cost-effectiveness of restorative materials for two and three surface restorations undertaken in Norway and found amalgam to be the most cost-effective, followed by composite and then gold.28 A similar analysis of cost-effectiveness in the UK also found amalgam to be the most cost-effective material.30 It was proposed that the cost-effectiveness of composites in particular was lower due to the shorter longevity and the higher cost of these restorations. A recent paper reported a systematic review of intra-coronal dental restorations in terms of their longevity and cost-effectiveness.15 It was noted that of the 30 economic studies identified, the majority were generally of poor quality, and the paper called for improved research in this area.

A more recent area of interest is that of implant retained prostheses. MacEntee and Walton looked at the costs associated with implant retained prostheses and conventional dentures.35 Jacobson et al. undertook one of the few utility based dental investigations in which implant retained prostheses and conventional dentures were compared using a rating scale method.36 They concluded that this was a reliable measure of patients' preferences and the implant group rated a successful implant-supported prosthesis as higher than a functional, fitting, aesthetic conventional denture, in spite of higher costs and longer periods of non-function.

There are relatively few cost-utility studies in the field of dentistry. A study in 1992 by Fyffe and Kay assessed the average utility values for four different 'tooth states' which it was hypothesised would have different values.39 They found that the highest mean utility values were for the restored tooth and lowest values for the decayed and painful posterior tooth. Values were obtained from both dentists and members of the general public and, perhaps not surprisingly, dentists gave higher utility values when compared with members of the general public. Downer and Moles also studied the influence of relevant factors on health gain from restorative treatment.40 O'Brien et al. undertook the only example which was found of utility analysis in orthodontics.41 They developed a TTO scale questionnaire using the aesthetic component of the Index of Treatment Need and found that patients seeking treatment gave lower utility values than those not wanting treatment. However, with the visual analogue scale there was no significant difference. It was proposed that this method could also be used as a method for predicting patient compliance.

In the field of oral medicine/oral pathology, Downer et al.42 used the Standard Gamble method to elicit the public's perceptions of different oral cancer states — precancer, small cancer and large cancer and found utility values of 0.92 for precancer, 0.88 for stage 1 cancer and 0.68 for stage 2 cancer. These values then allow the QALYs gained and the cost per QALY involved in the treatment of such lesions to be calculated.

The future

Only by improving research in economic evaluation and by improving planning and management systems will the health service progress. It is becoming increasingly obvious that demands for treatment can not be met and that choices need to be made. Alongside this, governments and third party payers have intensified their search for better value for money.

As more papers involving economic evaluation are seen in the literature, it is important that all those involved in the provision and purchasing of healthcare have a full understanding of the methods in current use. Also, an increasing number of clinicians are likely to be involved in this field of research and will be required to have a knowledge of the techniques.

There are many areas in dentistry which would benefit from clinical studies also incorporating some form of economic evaluation. Developments including new restorative materials, increased use of implants, aesthetic type dentistry (for example, complex crown and bridge work) and comparisons of adult and adolescent orthodontic treatment are all areas which could be studied.

7. Goold S D. Allocating health care: cost-utility analysis, informed democratic decision making, or the Veil of Ignorance? J Health
Development of a condition-specific quality of life measure for patients with dentofacial deformity: I. Reliability of the instrument


Abstract - The assessment of quality of life is becoming increasingly important in clinical research. Its importance in dentistry has been realised only relatively recently. Health-related quality of life is concerned with the aspects of quality of life that relate specifically to an individual’s health. This may be measured using two groups of instruments: (i) generic measures, which provide a summary of health-related quality of life and sometimes generate a single index measure of health or (ii) condition-specific measures, which focus on a particular condition, disease, population or problem and are potentially more responsive to small, but clinically important, changes in health.

Objectives: The aim of this study was to develop a condition-specific quality of life measure for those patients with severe dentofacial deformity who were requesting orthognathic treatment and to assess the reliability of this instrument.

Method: Instrument content was derived through a literature review and interviews with clinicians and patients. The resulting instrument was tested for internal consistency and test-retest reliability.

Results and conclusions: The instrument was found to divide into four clinically meaningful domains. Internal consistency and test-retest reliability were good. Patient acceptance of the questionnaire was also encouraging.

Key words: health related quality of life; orthognathic treatment

Susan J. Cunningham, Department of Orthodontics, Eastman Dental Institute, University College London, 256 Gray’s Inn Road, London WC1X 8LD, UK
Tel: +44 171 915 1000 ext. 1317
Fax: +44 171 915 1238
e-mail: s.cunningham@eastman.ucl.ac.uk
Submitted 3 November 1998; accepted 1 October 1999

The World Health Organization defines quality of life as “an individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad-ranging concept affected in a complex way by the person’s physical health, psychological state, level of independence, social relationships, and their relationships to salient features of their environment.” Quality of life (QOL) may also be defined as “a person's sense of well-being that stems from satisfaction or dissatisfaction with the areas of life that are important to him/her”. In spite of this pioneering work by WHO, it is only recently that this definition has been operationalised, and quality of life was introduced as a Medline heading as recently as 1975 (1).

Clinical research is usually concerned with the aspects of quality of life that relate specifically to an individual’s health – so-called “health-related quality of life” (HRQL). Within HRQL it is generally agreed that it is a multidimensional concept; however, there are two schools of thought. One group argues that HRQL is a multidimensional concept such that a health profile is produced, the argument being that measurements are only possible within a domain and not between domains. The other argument is that different attributes of
HRQL can be combined to produce a single index value (2). HRQL is increasingly used as an outcome in clinical trials, effectiveness research and research on quality of care. This has been facilitated by the growing body of evidence for the properties of HRQL measures (3).

Recent years have seen an increased interest in health-related quality of life for three main applications (2):

- Widening the range of outcome measures used in clinical trials and other evaluative studies
- Auditing the performance of groups or individuals
- Monitoring health levels in the community.

There are two groups of instruments which may be used within these applications:

- Generic measures, which provide a summary of HRQL and include the Short-Form 36-item Health Survey or SF-36 (4) and the EuroQol (2).
- Specific measures, which focus on a particular condition, disease, population or problem. Their narrow focus means that they are potentially more responsive to small, but clinically important, changes in health. A number of condition-specific quality of life measures have been developed in areas such as inflammatory bowel disease (5), rhinoconjunctivitis (6) and varicose veins (7).

Both approaches have their strengths and weaknesses and there are advantages to using both instruments in research (7, 8).

The importance of quality of life has seen widespread acceptance in medicine. However, oral health status has only recently been seen in these terms, and quality of life is a relatively new concept in dentistry. The need for a comprehensive approach to study the social and psychological impact of oral disease was first realised in the late 1980s. Reisine et al. (9) were amongst the first to realise the importance of oral health-related quality of life. The instrument was designed for use with an older age group and now exists in both the original format and a shortened form.

Patients presenting with severe dentofacial deformities may require a comprehensive orthodontic and surgical approach to their treatment (so-called orthognathic treatment). This treatment involves a course of fixed orthodontic appliances followed by surgery to correct the skeletal discrepancy. Treatment takes in the region of 2 years to complete. These patients tend to be in the younger age group and currently lacking is any instrument to determine changes in quality of life as a result of this mode of treatment.

The assessment of outcome of orthognathic treatment is a difficult concept in that the patient’s life is not extended by treatment and they are not made “well” by treatment in the conventional understanding of the word in health care. These are two of the features most often taken into account in quality of life research and make it a difficult area to investigate. However, it is generally accepted that the patients benefit from a psychological point of view with improved facial and dental appearance and the associated increased self-confidence which accompanies this. A longitudinal study of pre- and post-operative psychological characteristics of patients undergoing orthognathic treatment by Kiyak et al. (12, 13) confirmed these perceptions. They found high levels of satisfaction following surgery and patients perceived considerable improvements in their facial appearance and body image. The authors concluded that satisfaction following treatment was generally high, with patients viewing themselves more positively. Flanary et al. (14) also found high levels of post-operative satisfaction and a healthy psychological adjustment which led them to conclude that orthognathic treatment appeared to have a positive impact on quality of life.

The purpose of this study was to develop a condition-specific quality of life measure to be used in studies investigating the outcome of orthognathic treatment. The instrument was developed through...
a literature review and in-depth interviews with both clinicians and patients. The aim of this study was to develop a condition-specific quality of life measure using standard techniques of item derivation and testing (15). Such a measure should be capable of measuring changes in quality of life over time and be valid, reliable and responsive. This paper describes the development and testing of the reliability of the measure. The responsiveness and validity testing will be described in a later paper. Orthognathic treatment takes in the order of 2 years to complete and the full data analysis will only be possible at a later stage. The instrument is known as the Orthognathic Quality of Life Questionnaire (OQLQ).

Material and methods

Ethical approval was obtained from the Joint Research and Ethics Committees of all departments involved in the study.

The instrument was developed in three stages:

(i) Item generation
(ii) Item reduction
(iii) Testing of the instrument.

The first stage involved generating a list of items or questions, the content of which would reflect the impact of facial deformity on an individual's quality of life. Items were derived following a literature review and interviews with health professionals and patients. Unstructured interviews were conducted with 10 maxillofacial surgeons (senior registrar or consultant level) and 15 orthodontists (senior registrar or consultant level) from teaching hospitals and district general hospitals in the United Kingdom. The health-care professionals were a convenience sample selected on the basis that they were known to one of the authors (SJC) and could be readily interviewed by telephone. All clinicians interviewed were asked how they felt dentofacial deformities affected patients' quality of life and their responses were listed.

Interviews were also conducted with 15 patients who were about to start treatment and 10 patients who were in the early stages of pre-surgical orthodontic treatment. These individuals were randomly selected, based on the next 25 consecutive patients within the department who fulfilled the following criteria. Subjects were interviewed only if they were over 16 years of age, spoke English as a first language, did not have any congenital deformities (such as cleft lip and palate) and were embarking on a course of treatment involving both orthodontics and orthognathic surgery. All interviews were performed by one of the authors (SJC) and were undertaken away from the clinical area. Interviewees were asked to explain how dentofacial problems restricted their lifestyle or affected quality of life. All statements were listed along with those from the health professionals and when duplicate statements had been removed this produced a list of 42 items.

The second stage, item reduction, assessed the response frequency for each item. The list of items was administered to a group of 46 patients not involved in the item generation stage. Of these respondents, 30 completed the questionnaire prior to any treatment and the remaining 16 were undergoing pre-surgical orthodontic treatment. Respondents were asked to mark each item which was relevant in their initial decision to proceed with orthognathic treatment and to return the questionnaire in a sealed envelope. Those items which were selected most frequently were included in the measure. Items which were selected by 20% or less of the respondents were excluded, as they were considered poor at discriminating between patients (15). This method of item reduction produced a list of 22 statements which were used in the final measure.

Evaluation of the measure is currently taking place within a longitudinal study of orthognathic patients in three different orthodontic/maxillofacial units in the south of England. Eighty-eight patients were recruited for the study and the OQLQ was administered prior to any treatment and will also be administered after pre-surgical orthodontic treatment (prior to surgery) and at the end of treatment (6–8 weeks following removal of orthodontic appliances).

Respondents were provided with a questionnaire in which they were asked to mark the 22 statements based on a 4-point scale where 1 meant the issue covered in the statement bothered them a little, 4 meant it bothered them a lot and 2 and 3 were in between. An option of “not applicable” existed for those not affected by the issue covered in that statement (Appendix 1).

The results presented here are those based on the responses given at the start of treatment. A comparison of pre- and post-treatment responses will be presented at a later date. Of the 88 respondents, 24 were asked to complete exactly the same questionnaire 6 weeks following the first in order to assess test-retest reliability. This time interval was selected for the test-retest because the respondents
had not started treatment and therefore should have remained stable and, in addition, they were unlikely to remember how they answered the first time.

Statistical analysis
Statistical analysis was undertaken using SPSS for Windows. Analysis involved:
- Assessment of response frequencies to ensure that a substantial proportion of patients were not answering the same way for any individual item
- A principal component analysis (PCA) to look at the underlying dimensionality of the data and to determine whether the instrument was made up of subscales
- Assessment of internal consistency of the separate dimensions of the measure – this involved testing for homogeneity and measured correlations between items in each domain and in the instrument as a whole
- Test-retest to assess the level of agreement between the two sets of scores.

Results
In the item reduction stage, 46 patients were mailed questionnaires and 44 were returned giving a 96% response rate. As described in the previous section, the results of item reduction produced the 22-item questionnaire shown in Appendix 1.

Of the 88 respondents administered a questionnaire before treatment, 85 completed it giving a response rate of 97%.

Response frequencies were initially assessed to ensure that a substantial number of respondents were not answering the same way for any item, as such questions lack discriminatory power. No single item had a response frequency over 80%, therefore, it was not necessary to exclude any items at this stage.

The results of the PCA are shown in Table 1. There are four important components which ap-

<table>
<thead>
<tr>
<th>Component/item</th>
<th>Item-total correlation</th>
<th>Alpha coefficient for each component</th>
<th>Intra-class correlation coefficient (test-retest)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component 1 – social aspects of deformity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Cover mouth when meeting people</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Worry about meeting people</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Worry people will make hurtful comments</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Lack confidence socially</td>
<td>0.76</td>
<td>0.93</td>
<td>0.88** (P&lt;0.01)</td>
</tr>
<tr>
<td>19. Do not like smiling</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Get depressed about appearance</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Sometimes think people are staring</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Comments about appearance upset me</td>
<td>0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 2 – facial aesthetics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Self-conscious about appearance of my teeth</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Don’t like seeing side view of face (profile)</td>
<td>0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Dislike having photograph taken</td>
<td>0.82</td>
<td>0.86</td>
<td>0.92** (P&lt;0.01)</td>
</tr>
<tr>
<td>11. Dislike being seen on video</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Self-conscious about appearance</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 3 – function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Problems biting</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Problems chewing</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Avoid eating some foods</td>
<td>0.68</td>
<td>0.83</td>
<td>0.76** (P&lt;0.01)</td>
</tr>
<tr>
<td>5. Don’t like eating in public</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Pains in face/jaw</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Component 4 – awareness of facial deformity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Spend time studying face</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Spend time studying teeth</td>
<td>0.67</td>
<td>0.87</td>
<td>0.89** (P&lt;0.01)</td>
</tr>
<tr>
<td>12. Stare at people’s teeth</td>
<td>0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Stare at people’s faces</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
peared clinically meaningful. When these components were looked at individually, they divided into concerns regarding:

- First component – social aspects of deformity
- Second component – facial aesthetics
- Third component – oral function
- Fourth component – awareness of facial deformities.

The results of the internal consistency analysis are also shown in Table 1. A low item-total correlation would indicate that the item was unrelated to the items in the remainder of the scale. Only question 1 in the second and question 5 in the third components gave values below 0.5, all other items gave values between 0.52 and 0.88, which showed good relationships between the items. The alpha coefficients for the four components were high (between 0.83 and 0.93; significant P<0.01).

The test-retest analysis was performed by comparing the two sets of scores for each component using the Intra-class correlation coefficient. The results are shown in Table 1. All items were significantly correlated between the initial and the retest questionnaire.

Discussion

The methodology for developing the measure used several sources for item derivation including knowledge of the literature and unstructured interviews with clinicians and patients. This was followed by item reduction, where items which were chosen most frequently were included in the final measure, and exploratory PCA to determine dimensionality of the measure.

This study looked at a relatively new area for the investigation of health-related quality of life. Patients with severe dentofacial deformity have not been studied in this way previously. Although a relatively small number of patients were involved in the item reduction phase and in the testing of the instrument, these sample sizes were comparable with a number of other condition-specific quality of life studies (6, 16, 17). In addition, the results (particularly the exploratory PCA) gave clinically meaningful information which suggests an adequate sample size.

The items were generated using patients who were about to start treatment and patients were undergoing pre-surgical orthodontics. It would also have been interesting to include those who had completed treatment and to determine whether new quality of life items were discovered after patients had been treated. However, this was outside the remit of the present study. Another interesting area to look at is the relationship between demographic variables, such as age and socio-economic status, and quality of life. It is intended that these data will be presented alongside the validation data.

In developing condition-specific measures of HRQL, it has been stated that certain criteria should be fulfilled (5):

i. Summary scores should be amenable to statistical analysis.
ii. Repeated administration in stable patients must yield similar results.
iii. When even a small clinically important change has occurred, the questionnaire score should reflect it.
iv. The questionnaire should be valid.
v. The questionnaire should be relatively short and simple.

This study shows that criteria i, ii and v have already been fulfilled. The remaining criteria (iii and iv) will be investigated fully when the measure is administered to respondents at the end of pre-surgical orthodontics and again following completion of treatment.

This study showed that it was possible to produce a relatively short and simple list of items relating to physical and emotional function and that the scores produced could be statistically analysed. Analysis produced encouraging results. In addition, repeated administration in the test-retest group produced highly consistent results, suggesting the instrument is stable.

The measure has a number of potential uses in clinical practice and research. With increased emphasis on outcomes and quality assurance such a measure will produce important data. It may provide a basis for efficient allocation of resources and such a measure could be used in conjunction with economic evaluation to do so. It may also be used to compare the effectiveness of different procedures. For example, it may be used to compare treatment of mandibular retrognathia by distraction osteogenesis and by conventional orthognathic surgery or to compare outcomes of single jaw and bimaxillary surgery.

Acknowledgements

The authors would like to thank all the clinicians and patients who have given up their own time to help with this study.
Appendix 1

Final condition-specific measure

Please read the following statements carefully. In order to find out how important each of the statements is to you. Please circle 1, 2, 3, 4 or N/A where:

1 means it bothers you a little
4 means it bothers you a lot
2+3 lie between these statements
N/A means the statement does not apply to you or does not bother you

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References

14. Flanary CM, Branwell GM, VanSickles JE, Littlefield JH, Rugh AL. Impact of orthognathic surgery on normal and


Relationship between utility values and willingness to pay in patients undergoing orthognathic treatment

Susan J. Cunningham and Nigel P. Hunt

Orthodontic Department, Eastman Dental Institute, University College London

Objective To determine utility and willingness to pay (WTP) values for patients undergoing orthognathic treatment in a National Health Service hospital in the United Kingdom and to establish whether WTP values can be used as a measure of strength of preference. Participants Forty patients who were about to start orthognathic treatment were recruited over a 10-month period between January and October 1998. Outcome measures Utility values were obtained using the standard gamble method and WTP values were determined using a payment card method. The relationship between the two was assessed. Results The mean utility value was 0.73 and the mean WTP was £6,833. The correlation between these two measures was -0.46 (P<0.01). Conclusions The correlation between the utility and WTP values was in the expected direction thus suggesting that WTP may be used as a measure of strength of preference. Therefore, willingness to pay may be a useful technique to combine with cost utility analysis.

Key words: contingent valuation, orthognathic treatment; utility analysis

Introduction

Orthognathic treatment involves orthodontic treatment and surgery in order to treat those patients who present with dentofacial deformities outside the scope of routine orthodontic treatment. The treatment has a major impact on quality of life rather than life expectancy and these changes in quality of life may be measured in three main ways:

• Cost utility analysis – through the assessment of Quality Adjusted Life Years (QALYs) gained
• Contingent valuation – also known as willingness to pay
• Quality of life instruments – which may be generic (e.g. the Short Form-36) or condition specific

Cost-utility analysis

Cost-utility analysis (CUA) assigns a ratio of cost to benefit and allows efficient use of resources in a manner which is considered consistent with justice. Most people would agree that quality of life is an important consideration in the allocation of resources and the emphasis in this area has increased dramatically over the last few years. CUA allows comparison of a wide range of different interventions including those which are life saving and those which enhance quality of life (Drummond et al., 1996; Gerard et al., 1993). Utility based measures are usually expressed in terms of quality adjusted life years (QALYs). Health states are assigned a utility value between 0 and 1, where 0 is equivalent to death and 1 is equivalent to perfect health. QALYs are calculated by multiplying this utility value by the remaining life expectancy. Information from QALYs, along with costs, can then be used to guide resource allocation.

There are three accepted methods of obtaining utility values (Drummond et al., op.cit):

• Rating scale (RS)
The rating scale is easily understood but has no theoretical foundation. The measures from a rating scale tend to be referred to as values rather than utilities because the latter are measured under conditions of uncertainty. As the rating scale does not elicit cardinal utility measures, it should be used in conjunction with one of the other methods (standard gamble or time trade-off).

• The von Neumann and Morgenstem standard gamble (SG)
The standard gamble method is said to be valid because it is based on the von Neumann and Morgenstem utility theory for decision making under conditions of uncertainty (von Neumann and Morgenstern, 1953). This method has a number of advantages in that it reflects choice, the probabilities fulfil interval criteria and the units are readily understood. In addition, it reflects risk, something which is involved in any medical intervention. It has been accepted as the ‘gold standard’ for many years. Further details of this technique are included in the method (Figure 1).

• Time trade-off (TTO)
Torrance et al. (1972) originally developed the TTO because some respondents found the probabilities involved in the standard gamble difficult to understand. It was developed for use primarily in medical research. Subjects are offered Choice A (their current health state) for time t or Choice B, a shorter life span (time x) but in perfect health. The time in the healthy state (time x) is varied until an indifference point is reached, at which the utility value is recorded (Figure 2).
Contingent valuation

“political economy has to take as the measure of utility of an object the maximum sacrifice which each customer would be willing to make in order to acquire the object... the only real utility is that which people are willing to pay for.” (Dupuit, 1844, cited in Donaldson, 1993).

Contingent valuation includes the willingness to pay (WTP) and the willingness to accept (WTA) methods. Of these, WTP is the most commonly used. The WTP technique asks patients what they would be willing to pay for specific public goods. Contingent valuation was first used in the field of environmental issues but has more recently been introduced into health care. There has been a growing interest in the WTP method as a result of some of the limitations of QALYs. The technique of WTP can be used alongside other economic evaluation data in determining funding of certain medical interventions. This allows decision making to be based on accurate assessments of people’s values rather than what the policy makers believe is correct.

Willingness to pay studies may be performed using postal surveys, questionnaires or interviews. Recent WTP recommendations in the field of environmental benefits state that interviews should be used in preference to postal questionnaires and surveys (Arrow et al., 1993).

There are several possible methods for measuring WTP (Donaldson et al., 1995b; Mitchell and Carson, 1989):

1. Bidding game – This is the oldest and most widely used method. It has the disadvantage that the final amount may be influenced by the starting amount (starting bias).
2. Payment card method – This was developed as an alternative to the bidding game and avoids having to give a starting bid. A range of amounts is presented ranging from 0 to a high figure. However, the values must be realistic.
3. ‘Take it or leave it approach’ – A large number of prices are set and the respondent is asked if they would be prepared to pay one certain price. This has the disadvantage that there are those respondents who will automatically say yes (the so-called ‘yes bias’).
4. Open ended method – The respondent is asked how much they would be prepared to pay without being given any guide. Many respondents find this a difficult concept, particularly in Great Britain where patients are not generally familiar with payment for health care.
5. Split sample – Different methods can be used in the same study if the sample is of sufficient size.

In theory, WTP is superior to QALYs from a consumer perspective as there are no restrictions on the attributes which can be valued using WTP. However, there are several controversial issues associated with this method. Firstly, those who are affluent may be able to afford more in their WTP whilst those who have less money may be more readily compensated, thereby ignoring the considerations of distributive justice (Donaldson, op.cit). However, a study by Donaldson et al. (1995a) investigating preferences regarding obstetric care in Scotland found that the maximum WTP was not associated with indicators of ability to pay and that the results reflected the strength of the preference alone rather than being a reflection of the ability to pay. They concluded that the relationship between willingness and ability to pay need not count against the method.

Secondly, respondents may object to or misunderstand the concept and it is important that the respondents are told that this is a hypothetical situation. However, that in itself may lead to an overestimation of the preferred amount (Donaldson et al., 1995a). Many studies reveal respondents who object to the concept and register a ‘protest’ in that they feel that they should not be made to consider paying even in a hypothetical situation.

A third controversy is which method to use, whether to use the open-ended approach, closed-ended approach or payment card method (Donaldson et al., 1995b).

The aim of this study was to elicit utility values and willingness to pay values from patients who were about to commence orthognathic treatment. It was hypothesised that if WTP truly measures a strength of preference in the same way as utility values, then there should be a negative correlation between the two values. Therefore, respondents would be expected to be willing to pay increasing amounts of money as the utility value decreased.

Method

Subjects

The study took place in a National Health Hospital in the London area of the United Kingdom. Ethical approval
was obtained from the hospital's Joint Research and Ethics Committee. All respondents were advised that responses were confidential and would not affect their treatment in any way. Forty patients who presented with dentofacial deformities requesting combined orthodontic/maxillofacial treatment (orthognathic treatment) were recruited for the study. Patients were recruited over a ten month period between January and October 1998. All those who fulfilled the selection criteria were asked if they would participate. All subjects were over 16 years of age. Those who did not speak English sufficiently well to be interviewed, or those who presented with dentofacial deformities due to syndromes or cleft lip and palate, were excluded.

Method

Utility values were derived using the standard gamble method (Figure 1). This involved asking subjects to choose between two sets of health-related personal circumstances. In the first situation (Choice A) they were told to envisage that they would live with the dentofacial deformity for the remainder of their life. In the second situation (Choice B) they were asked to envisage a gamble situation where there were two possible outcomes, perfect health with no dentofacial deformity or death. The probability of perfect health was decreased incrementally and the probability of death increased incrementally. The value at which the subject was not willing to take any further risk was regarded as the indifference point and was recorded as the utility value. The more risk a respondent was prepared to accept, the lower the utility value.

The principles involved in the SG may be difficult for some respondents to understand and visual aids were used to make the concepts easier. For this reason, a card deck was constructed using laminated diagrams in which colour coded pie-chart segments represented the probabilities.

Willingness to pay values were ascertained using the payment card method. This method was selected as it has been suggested that it is more valid than an open-ended approach (Donaldson et al., 1995b). Cards were presented with values ranging from zero pounds sterling (£0) to fifteen thousand pounds sterling (£15,000) in one thousand pound (£1,000) increments. These values were chosen as the range was acceptable and incorporated within it the actual cost of the procedure. Respondents were asked how much they would be willing to pay to undergo treatment for the correction of their dentofacial deformity. They were advised to consider that the amount could be paid in instalments if required.

A number of the respondents were students and not financially self-sufficient. Greater than 50 per cent of these individuals did not know their household income and it was, therefore, decided not to relate willingness to pay values to household income. Donaldson et al. (1995b) noted that WTP was not associated with indicators of ability to pay and the results reflected the strength of the preference alone rather than being a reflection of ability to pay. The WTP method can, therefore, still be considered an acceptable measure of the strength of preference without consideration of the respondent's income. However, in an attempt to investigate bias in the method, postcodes were used to derive a Townsend deprivation index for each respondent (Townsend et al., 1988). This index consists of four population variables: percentage measures of local unemployment, car ownership, overcrowding and home ownership. The index may have a negative or positive value, with positive values indicating greater deprivation and negative values showing greater affluence. The relationship between this index and WTP was investigated.

Statistical Analysis

Data were analysed using the SPSS for Windows package. Both mean and median values were calculated, as well as the range of values and the standard error. Pearson correlation coefficients were calculated for the relationship between utility value and WTP and between the Townsend index and WTP.

Results

The results are presented in Tables 1 and 2.

Forty patients who had confirmed their decision to proceed with treatment agreed to participate in the study. All respondents who were approached agreed to take part and none refused to answer any part of the interview. Table 1 illustrates the demographic variables of the respondents. The female: male ratio was approximately 2:1 and the mean age 24 years.

Table 2 shows the utility and WTP values. The mean utility value was 0.73 (SE 0.039). The mean willingness to pay was £6,833 (SE £880).

There was a correlation of -0.46 between the utility and WTP values and this was significant at the 1 per cent level. This table also shows a correlation of -0.05 (P=0.76 ns) between the Townsend index and WTP.

Discussion

This study investigated utility and willingness-to-pay for those individuals requesting orthognathic treatment for the correction of dentofacial deformity. The study does include a relatively small number of patients but nevertheless gives a good introduction to the possibility of using these techniques in dental research. Recruitment of orthognathic patients is recognised as being difficult due to the relatively small number of individuals undergoing this form of treatment. All patients who commenced treatment over the study period were included.

This study confirmed that both the utility and the willingness to pay methods were acceptable to patients requesting treatment for dentofacial deformity and it is now intended that larger numbers of patients should be investigated using these techniques. The SG technique appeared to be understood by the respondents, even though the health state in question was not life threatening, and

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (years)</th>
<th>Mean</th>
<th>(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female</td>
<td>13</td>
<td>27</td>
</tr>
</tbody>
</table>
Table 2: Statistical analysis of utility and willingness-to-pay values including the correlation coefficient for the relationship between the two values.

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Range</th>
<th>Mean</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health state utility</td>
<td>0.85</td>
<td>0.05-0.95</td>
<td>0.73</td>
<td>0.039</td>
</tr>
<tr>
<td>Willingness to pay</td>
<td>£5,000</td>
<td>£0-15,000</td>
<td>£6,833</td>
<td>£880</td>
</tr>
</tbody>
</table>

Correlation for utility vs WTP value

-0.46 (P<0.01)

Correlation for Townsend index vs WTP value

-0.05 (P=0.76 ns)

this method allows dentofacial deformity and the oral cavity to be considered in the overall concept of the individual's health.

However, it must be noted that the standard errors associated with both methods were quite high and results obtained using them must be treated with a certain amount of caution. The results show variations between subjects for both utility and WTP values. There was no significant relationship between age, gender or malocclusion and the utility/WTP values. It is more likely that these variations can be explained in terms of the different reactions by individuals to their dentofacial discrepancy. This is supported by the fact that it is not unusual to see a patient who presents with a mild deformity and is very distressed by it or, equally, to see someone who presents with a severe dentofacial problem but declines treatment.

The correlation between utility value and WTP was in the expected direction (-0.46 P<0.01) which means that those individuals who gave the lowest utility values (e.g. were prepared to risk most for treatment) were also willing to pay the most. This suggests that willingness to pay can be used as a measure of strength of preference. O'Brien and Viramontes (1994) also found a correlation of -0.46 between the SG and WTP in their study of patients with chronic lung disease.

The relationship between Townsend deprivation index and WTP was also in the expected direction (e.g. as social deprivation increases the WTP is less) but this was not significant (P=0.76). In this study, the WTP was therefore not influenced by ability to pay which, again, suggests that WTP can be used as a measure of strength of preference. It is recognised that the Townsend index is not without flaws because there is no individual measure of social class; the index gives only an area measure of residency and does not pertain to the individual. However, this was the best indication available to the authors.

Comparison with other studies is very difficult. There are still a limited number of WTP studies in health care and most of these have been undertaken in the United States or in other countries where payment for medical care is the norm. This makes the concept of WTP easier for respondents to understand. In addition, some studies have estimated WTP as a percentage of household income, an area which the authors were unable to investigate in this study. For example, Thompson (1986) noted that individuals were willing to pay up to 22 per cent of their household income for a cure for rheumatoid arthritis.

The results from this study can also be used to estimate the willingness to pay per quality adjusted life year gained (QALY gained). The QALYs gained as a result of treatment may be calculated using the formula:

\[
\text{QALYs gained} = \text{improvement in utility as a result of treatment} \times \text{remaining life expectancy}
\]

If a post treatment utility of 1.0 is assumed

\[
\text{QALYs gained} = 0.27 \times 50 = 13.5
\]

The mean willingness to pay per quality adjusted life year gained is then calculated as £506. Again, this value is difficult to compare with the few studies which have quoted WTP per QALY gained as few values are quoted in UK pounds. For example, Olsen and Donaldson (1998) quoted mean values in Norwegian kroner and many studies are quoted in US dollars (Blumenschein and Johannesson, 1998; O'Brien and Viramontes, op. cit.).

The use of WTP remains a controversial issue. However this study suggests that it may be usefully combined with utility analysis when looking at strength of preference for health care intervention. Larger sample sizes would be of value for further investigation. It would also be useful to compare different forms of health care intervention in the same study. For example, comparison could be made of the willingness to pay for orthognathic treatment and adult orthodontic treatment.

Acknowledgements: The authors wish to acknowledge the valuable assistance of Professor George Torrance of McMaster University, Ontario and Dr France Sassi of the London School of Economics for their advice regarding utility measurement. Dr Emma McIntosh from the Health Services Research Unit at the University of Aberdeen for her advice regarding contingent evaluation and Dr Mark Gilthorpe for his assistance with the use of the Townsend index.

References


Are orthognathic patients different?

Susan J. Cunningham*, Mark S. Gilthorpe** and Nigel P. Hunt*
*Department of Orthodontics and **Biostatistics Unit, Eastman Dental Institute, University College
London, London, UK

SUMMARY This questionnaire-based study investigated the psychological profile of orthognathic patients prior to starting treatment and compared the findings with a control group of non-patients. Comparison of the data used multivariate multiple regression analysis where outcome variables and independent variables were studied simultaneously. Some differences were found in the psychological profile of the orthognathic patient. They displayed higher levels of state anxiety ($P < 0.001$), higher numbers of individuals in their social support network ($P < 0.05$), and lower body image and facial body image ($P < 0.001$). Self-esteem was also found to be lower, but only at borderline levels of significance ($P = 0.052$).

Introduction

As orthognathic treatment has become more readily available and more socially acceptable, the demand for treatment has increased considerably. Interest in the psychological aspects of combined orthodontic and surgical treatment has also increased, although the psychological issues are still not fully understood.

A review of the literature suggests a very high level of satisfaction and fewer post-operative problems following orthognathic surgery in contrast with studies involving cosmetic surgery patients, which in this context includes rhinoplasty, breast augmentation, and breast reduction (Heldt et al., 1982). A number of reasons have been proposed for these findings, including a lower level of psychiatric disorders in those seeking orthognathic surgery (Heldt et al., 1982). This, in turn, may lead to greater levels of post-surgical satisfaction.

Several authors have investigated the psychological profile of patients undergoing orthognathic treatment. Generally, these studies have been undertaken prior to surgery, rather than at the start of treatment (i.e. prior to the pre-surgical orthodontics). Flanary et al. (1990) investigated psychological adjustment and self-concept in 61 patients pre- and post-surgery. The patients were found to be relatively healthy and well adjusted, both pre- and post-operatively. However, significant improvements after surgery were noted in the following areas: psychoses, neuroses, personality disorders, and integration. A positive effect was also observed in all subscales of self-concept. Cunningham et al. (1996) looked at levels of depression, anxiety, and self-esteem before and after surgery, but found no significant difference at the two time periods.

The aim of this study was to assess the psychological profile of orthognathic patients prior to any active treatment and to compare these values with a control group. It is reasonable to assume that some changes may take place in the psychological profile during the pre-surgical orthodontic phase and it is for this reason that the patients were assessed prior to this treatment.

Materials and methods

A questionnaire was developed which assessed various aspects of the respondent’s psychological profile. Areas were selected which it was felt may play a part in the individual’s responses during treatment and also their satisfaction with the outcomes of treatment:

1. anxiety (two measures: one each for state and trait anxiety);
2. depression;


(3) perception of social support (two measures: the number of individuals in the support network and the respondent's satisfaction with their support network);
(4) self-esteem;
(5) body image (two measures: one for general body image and one for facial body image).

Components of the questionnaire

Anxiety. Anxiety was assessed using the State-Trait Anxiety Inventory, which was developed during the 1960s and revised in 1983 (Spielberger et al., 1983). The questionnaire was developed from a large pool of statements and was extensively tested on college students. Although there are a number of instruments to measure anxiety, this is one of the most widely used in psychological and clinical research. Anxiety is generally viewed as taking two forms—the state form (transitory feelings of fear/worry) and the trait form (the stable tendency to respond anxiously to stressful situations). The instrument consists of 40 items, 20 of which measure state anxiety and 20 that measure trait anxiety. Each item is scored on a 4-point scale. The use of the instrument in this study ensured that it was used on a population with a similar age range as those for which it was developed.

Depression. The Beck Depression Inventory (BDI) was used to assess depression (Beck et al., 1988). The BDI covers a wide range of items from sadness and sense of failure to self-dislike, social withdrawal, and indecisiveness. There are 21 items with scores ranging from 0 to 3. The BDI has been shown to have good psychometric properties and to be sensitive to change (Beck et al., 1988).

Social support. Social support was included because there is now a widely held belief that social support affects outcomes of health intervention (Cohen and Wills, 1985). Social support is commonly seen as a resource, which can be mobilized in times of crisis to protect against health threats. Social support is a broad term—referring to social, emotional, and other supports that are provided by an individual’s social contacts (Cohen and Wills, 1985). A study by Holman et al. (1995) showed that general social support was important immediately post-surgery for orthognathic patients and the reactions of individuals in the social support network were important in the later post-operative periods as the patient integrates the changes into their body image.

There are a number of social support measures available, but these are greatly reduced if those that are lengthy and time-consuming to complete are excluded. The Short Form Social Support Questionnaire (Sarason et al., 1987) was selected for use as it is relatively quick and easy to complete. It is a 6-item questionnaire, which measures perceived levels of support and satisfaction with this support.

Self-esteem. The concept of self-esteem or perceived self-worth is part of the wider construct of self-concept. It is a personal resource that may moderate the effects of conditions or events, including conditions such as disfigurement. Self-esteem is a dimension of life satisfaction and is an important part of general assessment of life (Andrews and Withey, 1976).

This study utilized the Rosenberg Self-Esteem (RSE) Scale, which was developed by Rosenberg (1965) for a study of students in New York. It is a 10-item scale with a 4-point scale of agreement with half the items being expressions of positive self-esteem and half negative.

Body image. Body image may be a source of distress for individuals of all ages. ‘Body image’ is a concept which has become more of an issue with the large amount of research in recent years focusing on those individuals with eating disorders (Lautenbacher et al., 1993; Nelson and Gidycz, 1993). Body image is considered as a complex psychological concept related to the mental representation of self and it is therefore reasonable to assume that a change in body image resulting from surgical procedures must have an impact. It is possible that this may be influenced by the surgery itself and the response of others to the surgical outcome. Body image is also thought to be related to self-esteem. A body or facial defect may affect self-esteem directly by negative feedback from others or indirectly by affecting the ability to master
ARE ORTHOGNATHIC PATIENTS DIFFERENT?

A revised version of Secord and Jourard's Body Cathexis Scale with a specific section asking about facial features was selected for use in the study (Kiyak et al., 1986; Finlay et al., 1995). The whole scale was analysed for body image and 10 items concentrating on facial features were also looked at in isolation.

Subjects

This study was part of a larger longitudinal investigation into the psychological profile of orthognathic patients during treatment. Ethical approval was received from the Joint Research and Ethics Committees of all centres involved. All participants signed a consent form, and it was stressed that responses were confidential and would not affect treatment in any way.

The questionnaire was distributed to 84 patients in three different centres. All patients had attended the Joint Orthodontic Maxillofacial Clinic with a view to starting orthognathic treatment. Patients were asked to complete the questionnaire at home as it took in the region of 30-40 minutes to complete. They were provided with a stamped addressed envelope and asked to return the questionnaire within 1 week following the appointment. Respondents who did not reply within 2 weeks were contacted by telephone. All patients approached agreed to take part in the study and only three failed to return the questionnaire, leaving a total number of 81 respondents (response rate 96 per cent).

A control group was recruited from local colleges and offices and was matched by age, sex, and ethnic group as closely as possible with the experimental group. A total of 106 individuals were approached and 95 questionnaires were returned (response rate 90 per cent).

Components of the questionnaire that were not fully completed or were incorrectly completed were excluded from the analysis. Data was complete for the Beck Depression Inventory, but all other scales had some data missing. One respondent failed to answer for state anxiety, three for trait anxiety, 10 failed to complete the social support questions, and self-esteem and body image questions were omitted by four respondents. The non-completion of the social support measure was felt to reflect the additional effort required for its completion.

Statistical analysis

The data were analysed using the MLwiN regression analysis programme (Multilevel Models Project, Institute of Education, University of London, UK). In view of the fact that each control group respondent was not individually matched (one-for-one) for age, gender, and ethnic group with a comparable respondent in the experimental group, multiple regression analysis was undertaken where these patient characteristics were considered simultaneously for their impact on the eight-outcome measures. The outcome measures were analysed simultaneously using multivariate techniques. This method allowed any underlying demographic differences to be accounted for whilst evaluating the impact of dentofacial deformity (i.e. whether the patient was in the experimental or control group).

Age was treated as a continuous variable and the mean age of all 176 respondents was subtracted from individual age in order to centre the variable. This provided a meaningful interpretation of the regression coefficient when the age variable adopts the value of zero. Gender and dentofacial groups were binary variables, and were therefore allocated values of 1 or 0 for male or female, and experimental or control group. Ethnic group was introduced into the regression analysis separately and each group was contrasted with the largest group (white). The ethnic groups recorded were white, South Asian (Bangladeshi, Bengali, Indian, and Pakistani), black (black African, black Afro-Caribbean and black other), and other (including Moroccan, Cypriot and Persian).

Each outcome variable was assessed for normality, as this distributional property was required prior to undertaking regression analysis. Values that were skewed or kurtosed were transformed to produce a more normal distribution using functions of the natural logarithm.
All outcome measures were then standardized in order to convert deviations from the mean to units of standard deviations. This permitted comparison across all outcome measures of the relative differences in psychometric score between the two dentofacial groups.

The multivariate method yields eight simultaneous regression equations:

\[ y^{(k)} = \beta_0^{(k)} + \beta_1^{(k)} \times \text{facial group} + \beta_2^{(k)} \times \text{gender} + \beta_3^{(k)} \times \text{centred age} + \beta_4^{(k)} \times \text{ethnic 1} + \beta_5^{(k)} \times \text{ethnic 2} + \beta_6^{(k)} \times \text{ethnic 3} \]  

where \( k = 1, \ldots, 8 \) and each outcome variable is normally distributed.

Parameters in equation 1 that were not significant at the 95 per cent level were removed from the final regression model. Correlations between the outcome measures were also studied.

Results

The results are shown in Tables 1-4.

Table 1 illustrates the values for the experimental and control groups prior to transformation and standardization. Only the first five psychometric scales required logarithmic adjustment prior to standardization.

Table 1  Values prior to transformation and standardization.

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>State anxiety (range 0–80)</td>
<td>20</td>
<td>77</td>
</tr>
<tr>
<td>Trait anxiety (range 0–80)</td>
<td>25</td>
<td>78</td>
</tr>
<tr>
<td>Depression (range 0–63)</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td>Social support (number) (range 0–9)</td>
<td>0.3</td>
<td>9.0</td>
</tr>
<tr>
<td>Social support (satisfaction) (range 0–6)</td>
<td>1.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Self-esteem (range 10–40)</td>
<td>11.0</td>
<td>37.0</td>
</tr>
<tr>
<td>Body image (range 27–135)</td>
<td>34.0</td>
<td>123.0</td>
</tr>
<tr>
<td>Facial body image (range 10–50)</td>
<td>16.0</td>
<td>47.0</td>
</tr>
</tbody>
</table>

NB: Scoring for all scales
- Anxiety  Increasing anxiety = higher scores
- Depression Increasing depression = higher scores
- Social support Increasing social support = higher scores
- Self-esteem Increasing self-esteem = lower scores
- Body image Increasing body image = lower scores
Table 2  Results of multivariate analysis (confidence intervals in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>Constant*</th>
<th>Group</th>
<th>Gender</th>
<th>Age (per 10 years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>35.6 (34.0, 37.3)</td>
<td>3.2 (1.2, 5.4)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>40.8 (39.2, 42.5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Depression</td>
<td>5.8 (5.0, 6.7)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Social support (number)</td>
<td>2.5 (1.3, 3.9)</td>
<td>0.7 (0.1, 1.3)</td>
<td>-0.6 (-1.0, -0.1)</td>
<td>-</td>
</tr>
<tr>
<td>Social support (satisfaction)</td>
<td>4.7 (5.3, 3.8)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>20.3 (19.3, 21.4)</td>
<td>1.1 (-0.0, 2.1)</td>
<td>-2.7 (-3.8, -1.6)</td>
<td>-</td>
</tr>
<tr>
<td>Body image</td>
<td>74.5 (70.7, 78.4)</td>
<td>9.3 (4.8, 13.8)</td>
<td>-8.5 (-13.1, -4.0)</td>
<td>-</td>
</tr>
<tr>
<td>Facial body image</td>
<td>27.3 (24.6, 30.0)</td>
<td>8.9 (7.2, 10.6)</td>
<td>-1.9 (-3.6, -0.2)</td>
<td>1.3 (0.3, 2.3)</td>
</tr>
</tbody>
</table>

Table 3  Significant parameters for each outcome variable.

<table>
<thead>
<tr>
<th>Significant parameters</th>
<th>Group</th>
<th>Gender</th>
<th>Age</th>
<th>Ethnic group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety (state)</td>
<td>***</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Anxiety (trait)</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Depression</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Social support (number)</td>
<td>*</td>
<td>*</td>
<td>NS</td>
<td>**</td>
</tr>
<tr>
<td>Social support (satisfaction)</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>NS</td>
<td>***</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Body image</td>
<td>NS</td>
<td>***</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Facial body image</td>
<td>***</td>
<td>NS</td>
<td>**</td>
<td>*</td>
</tr>
</tbody>
</table>

*Baseline values are those for white female controls aged 22 years.

Gender was significant for four outcome variables—number of individuals in the social support network, self-esteem, body image, and facial body image. With respect to females, males were likely to report fewer individuals for social support (mean difference of 0.6 people). Males also reported higher self-esteem (i.e. lower psychometric scores), and greater satisfaction with their body image and facial body image (also indicated by lower psychometric scores).

There were some significant differences between white and minority ethnic respondents, even though not all individual minority ethnic group terms were significant. Ethnic group was
The significant parameters for each outcome measure are shown in Table 3. Group was significant for state anxiety, social support (number), body image, and facial body image with borderline significance for self-esteem.

Table 4 gives the pre-model (or crude) outcome variable correlations. These did not alter significantly after modelling (not presented). The most notable correlations were for trait and state anxiety (0.73), depression and trait anxiety (0.74), self-esteem and trait anxiety (0.71), self-esteem and depression (0.62), and body image and facial body image (0.82).

Discussion
The results show that there are several parameters that affect the eight outcome variables. The only significant parameter affecting the difference in the state anxiety scores was whether the respondent was in the experimental or control group. As state anxiety is defined as transitory feelings of fear or worry, it is perhaps not surprising that the experimental group achieved a higher score in this variable. In contrast, in the dentofacial group there was no significant effect with respect to trait anxiety—the stable tendency to respond anxiously to stressful situations. Thus, the experimental and control groups did not differ fundamentally in their stable tendency to respond to anxious situations, although they exhibited different transitory levels of anxiety. This supports the concept that the two forms of anxiety should be assessed separately.

The dentofacial group was not significant for the BDI. This is encouraging in that orthognathic patients do not show higher levels of depression at the start of treatment. Therefore, the depression experienced by some patients following surgery is likely to be a reactive depression as a result of the treatment, rather than endogenous depression. The report of higher numbers in the support network of the experimental group may be relevant during treatment, even though satisfaction with social support did not show significant differences. Holman et al. (1995) noted the importance of social support, particularly in the post-operative period and this may influence satisfaction following treatment.

Although the dentofacial group only showed borderline significance for self-esteem values, this is an interesting area that could be studied further, possibly with a larger sample size. The

---

Table 4 Correlations for the model.

<table>
<thead>
<tr>
<th></th>
<th>State anxiety</th>
<th>Trait anxiety</th>
<th>Depression</th>
<th>Social support (number)</th>
<th>Social support (satisfaction)</th>
<th>Self-esteem</th>
<th>Body image</th>
<th>Facial body image</th>
</tr>
</thead>
<tbody>
<tr>
<td>State anxiety</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trait anxiety</td>
<td>0.73</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>0.56</td>
<td>0.74</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support (number)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support (satisfaction)</td>
<td>-0.27</td>
<td>-0.29</td>
<td>-0.22</td>
<td>0.21</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-esteem</td>
<td>0.54</td>
<td>0.71</td>
<td>0.62</td>
<td>-</td>
<td>0.35</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body image</td>
<td>0.39</td>
<td>0.49</td>
<td>0.46</td>
<td>-</td>
<td>0.23</td>
<td>0.58</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Facial body image</td>
<td>0.47</td>
<td>0.44</td>
<td>0.41</td>
<td>-</td>
<td>0.26</td>
<td>0.51</td>
<td>0.82</td>
<td>1.00</td>
</tr>
</tbody>
</table>
parameters that affect self-esteem are still largely unknown. The dentofacial group findings regarding body image are not surprising, as it is reasonable to assume that a patient who is prepared to go through extensive treatment to change their appearance probably has lower body image, particularly with respect to their facial features. The facial body image variable revealed the largest number of significant independent variables in the regression model. Satisfaction with facial body image decreased with age. It may therefore be interesting to study a group of older patients requesting orthognathic surgery and determine whether this is generally the case. This may also have some bearing on the success of outcome when operating on older patients.

The results regarding gender showed that males reported fewer individuals in their social support network, but interestingly satisfaction with support was not significantly different. Females showed lower self-esteem and reduced satisfaction with body image. Kiyak et al. (1981) reported that aesthetic improvement was important for both males and females, but it would still seem likely that females are starting from a lower baseline value than males.

Ethnic group was significant with respect to the number of individuals in the social support network, with all minority ethnic groups reporting a higher number of individuals compared with the white group. This is important because it may be that white patients need additional support during treatment and the clinician should be particularly vigilant. Satisfaction with social support was significant, with the South Asian group having the major effect. It is of interest that all minority ethnic groups reported higher scores on the facial body image scale than whites, indicating that facial appearance may be more important in the white culture.

The reason multivariate multiple regression analysis was undertaken was two-fold: to optimize the modelling for each psychometric scale through consideration of all eight variables simultaneously and to account for any residual demographic differences between the experimental and control groups, since respondents had not been matched one-for-one. Therefore, independent demographic variables that were found to be significant may have resulted from poor matching. Alternatively, differences might be inherent demographic contrasts measured by the associated psychometric scale. However, age was shown to be significant only for facial body image, and gender differences reflected anticipated differences between males and females, both of which suggested good matching for these variables. It was anticipated that matching would be most difficult for ethnic groups. However, there is little known regarding ethnic group differences among the psychometric measures investigated, and it is likely that there was a combination of both poorer matching and some cultural variability among the psychometric scales.

Correlations in the null model differed very little from those in the final model. In conjunction with the knowledge that matching was generally good, it is therefore reasonable to conclude that the chosen psychometric variables are demographically and culturally stable. That is the measures used were robust across the age, gender, and ethnic group composition of the target study groups. This is encouraging as one of the criticisms sometimes levelled at psychometric tests is that they are not culturally diverse.

Even though state and trait anxiety were measuring different aspects of anxiety, the high correlation between state and trait anxiety is perhaps not surprising. Neither is the correlation between body image and facial body image, since the latter is a sub-component of the body image scale. However, the correlations between self-esteem and trait anxiety (0.71), and self-esteem and depression (0.62) were interesting. As trait anxiety or depression increased, self-esteem was found to decrease and vice versa.

There are a number of negative correlations relating to social support. For the final model, as social support (number and satisfaction) increased, state and trait anxiety, and depression were reduced. This is an encouraging finding for surgical patients in that if their social support network can be encouraged to play a part in the treatment, they may show less anxiety and depression, both during and after treatment. As social support (numbers) increased, self-esteem, body image, and facial body image values
reduced (correlations of -0.28, -0.09, and 0.16). However, it must be borne in mind that all the negative correlations were very low.

Conclusions
Multivariate multiple regression analysis has a number of advantages. It allows comparison of experimental and control groups, taking account of demographic composition (i.e. without the need for one-for-one matching—a process which is recognized as being both difficult and time consuming) and can model differences in several psychometric scales simultaneously.

This study showed that there were some differences in the psychological profile of orthognathic patients with respect to control group respondents. The orthognathic patient was shown to be more likely to display higher levels of state anxiety, greater social support (number), lower body image, and facial body image. It is also possible that self-esteem is lower in the experimental group, although the significance was only borderline.

Acknowledgements
The authors wish to thank all the consultants who allowed access to their patients, as well as all the patients who completed the questionnaires.

Address for correspondence
S. J. Cunningham
Department of Orthodontics
Eastman Dental Institute
University of London
256 Gray's Inn Road
London WC1X 8LD, UK

References
Cohen S, Wills T A 1985 Stress, social support and the buffering hypothesis. Psychological Bulletin 98: 310-357
Rosenberg, M 1965 Society and the adolescent self image. Princetown University Press, Princetown
A comparison of health state utilities for dentofacial deformity as derived from patients and members of the general public

Susan J. Cunningham and Nigel P. Hunt
Orthodontic Department, Eastman Dental Institute, University College London, UK

SUMMARY The cost-utility approach is a method of economic evaluation, which assigns a ratio of cost to benefit, based on utility values of the health state in question. It allows efficient use of health care resources and is a useful method in that it permits comparison of a wide range of medical interventions, including those which are life saving and those that improve quality of life. This study obtained utility values for dentofacial deformity from orthognathic patients and members of the general public using three recognized methods—rating scale (RS), standard gamble (SG), and time trade-off (TTO). There were no significant differences between the utility values for the two groups of respondents. Method agreement between the TTO and the SG (the ‘gold standard’) was better than that between the RS and SG. In addition, the SG and TTO were found to have greater repeatability than the RS.

Introduction

Demand for health care has always exceeded the available resources. This raises the question of how to decide where the money should be most appropriately allocated. Economic evaluation is an accepted tool for the appraisal of health care programmes, and can be defined as the comparative analysis of alternative courses of action in terms of both their costs and consequences (Drummond et al., 1996). Evaluation is dependent upon the quality of underlying medical evidence and, because of this, clinical trials are increasingly viewed as a natural vehicle for economic analysis (Drummond and Davies, 1991). When ‘rationing’ of resources becomes necessary, some procedure has to be set to allow the most appropriate allocation. This was the basis for the introduction of cost-utility analysis (CUA) which assigns a ratio of cost to benefit and allows efficient use of resources in a manner that is considered consistent with justice.

Utility analysis

Most people would agree that quality of life is an important consideration in the allocation of resources and the emphasis in this area has increased dramatically over the last few years. CUA allows comparison of a wide range of different interventions, including those which are life saving and life enhancing (Gerard et al., 1993). Utility-based measures are usually expressed in terms of quality adjusted life years (QALYs). Health states are assigned a utility value between 0 and 1, where 0 is equivalent to death and 1 is equivalent to perfect health. QALYs are calculated by multiplying this utility value by the life expectancy. Information from QALYs, along with costs, can then be used to guide resource allocation. Utility values are therefore not the same as QALYs, but are used as a weighting factor to adjust the remaining life years for the quality of life that will be experienced (Bakker and van der Linden, 1995).
The main disadvantages of this method are that the interviews are labour intensive and, in addition, utility scores are not always easily interpreted. Despite these disadvantages, the technique remains popular and is likely to do so until a more appropriate method becomes available.

Existing measurement techniques

It has been widely assumed that the different measuring techniques produce comparable results, but this has not been investigated widely by practitioners of CUA. The following three methods are the most commonly used (Drummond et al., 1996).

Rating scale (RS). This is easily understood, but has no theoretical foundation. The measures from the RS tend to be referred to as values, rather than utilities because the latter are measured under conditions of uncertainty. Because the RS does not elicit valid cardinal utility measures, it should be used in conjunction with one of the other methods (SG or TTO).

The von Neumann and Morgenstern Standard Gamble. The standard gamble (SG) method is said to be valid because it is based on the utility theory for decisions under uncertainty (von Neumann and Morgenstern, 1953). This method has a number of advantages in that it reflects choice, the probabilities fulfil interval criteria, and the units are readily understood. In addition, it reflects risk, something that is involved in any medical intervention. It has been accepted as the ‘gold standard’ for many years.

Time Trade-Off. Torrance et al. (1972) originally developed the Time Trade-Off (TTO) for use in medical research because some respondents found the probabilities involved in the SG difficult to understand.

Cost utility and dentistry

There are few cost-utility studies in the field of dentistry. A recent Medline search found only 18 papers in this area published between 1980 and 1998. Fyffe and Kay (1992) assessed the average utility values for four different ‘tooth states’, which it was hypothesized would have different values. They found that the highest mean utility values were for the restored tooth and lowest values for the decayed and painful posterior tooth. Values were obtained from both dentists and members of the general public and, perhaps not surprisingly, dentists gave higher utility values.

Downer et al. (1997) used a convenience sample to elicit the public’s perceptions of different oral cancer states (pre-cancer, small cancer, and large cancer). They used a SG questionnaire and found utility values of 0.92 for pre-cancer, 0.88 for stage 1 cancer, and 0.68 for stage 2 cancer. There were statistically significant differences between all three values with the order of magnitude in the expected direction.

The future in the field of orthodontics

The lack of papers in this field is somewhat surprising in view of the fact that this method is useful for interventions, which improve quality of life, as many dental interventions do. The utility approach is particularly appropriate for interventions such as orthodontics and orthognathic treatment where the main argument for undertaking treatment is improved quality of life. In the future, it seems likely that healthcare providers (such as the National Health Service or other insurance agencies) may ask for proof that procedures such as orthognathic treatment should be publicly funded. Utility analysis is one of the areas, which may then prove useful. The utility value is essentially a measure of how good or bad the individual perceives the health state in question. If patients with dentofacial deformity consistently provide high utility values (in the region of 0.95–1.0) then it may be questionable whether the service should be provided. However, if pre-treatment values are consistently lower than this and there is a significant change in the perceived utility value following treatment, then this provides a very good argument for the continued provision of the service.

This study aimed to obtain utility values for the pre-treatment health state (e.g. start of
utility values for dentofacial deformity

337

treatment) for patients requesting combined orthodontic/maxillofacial surgery to correct dentofacial abnormalities.

Subjects and methods

The control group comprised a convenience sample of 55 adults (19 males and 36 females, mean age 24.0 years) from local colleges and businesses. All subjects were fit and healthy and none had any personal involvement in orthognathic procedures. The experimental group included 40 patients (13 males and 27 females, mean age 24.8 years) who presented with dentofacial deformities requesting orthognathic treatment in order to correct the problem. There was no significant difference in the gender or age distribution for the two groups.

The three methods already described (RS, SG, and TTO) were used to establish utility values for pre-treatment dentofacial appearance and function. One of the most important factors when using utility values derived from the general public is that they must understand fully the health state being described. The study aimed to achieve this using verbal explanations and black and white photographs of orthognathic patients. Photographic records of four patients were selected (male and female; Class II and Class III malocclusions) and laminated on card. The process of orthognathic treatment was also described to each respondent.

Rating scale

As described by Feeny and Torrance (1989) the RS was a vertical and calibrated visual analogue scale (0–100). The choice of a 0–100 dead-healthy scale was chosen such that the data obtained could be compared with a wide number of other medical interventions. Initially the ends of the scale were 'anchored'. This was achieved using two health state descriptions:

1. Perfect health state—respondents were 'forced' to assign perfect health to a scale value of 100.
2. Death—respondents were told that if death was the worst health state they could imagine, it should be placed at 0. However, if worse health states could be imagined (for example, being in constant pain), death was placed at an intermediate point (d) on the scale.

The scale was then used to record a preference value (x) for the health state under test. If death was marked at a point (d) rather than at 0, the preference values of the health state took this into account using Equation 1.

\[
\text{Value for health state} = x - \frac{d}{100} - d 
\]  

(eqn 1)

Standard gamble

Subjects were asked to choose between two sets of health-related personal circumstances. In the first situation (Choice A), they were told to envisage that they would live with the dentofacial deformity for the remainder of their life. In the second situation (Choice B), they were asked to envisage a gamble situation where there were two possible outcomes, perfect health with no dentofacial deformity or immediate death (Figure 1). The probability of perfect health was decreased incrementally and the probability of death increased incrementally. The value at which the subject was not willing to take any further risk of immediate death was regarded as the indifference point and was recorded as the utility value. The more risk a respondent was prepared to accept, the lower the utility value.

\[
\begin{align*}
\text{Probability } p \text{ (perfect health)} \\
\text{Choice B} \\
\text{Probability } (1-p) \text{ (death)} \\
\text{Choice A} \\
\text{State I (dento-facial deformity)}
\end{align*}
\]

Figure 1 The standard gamble technique.
CHOICE A
100% chance of remaining in present state for remainder of life (e.g. with dentofacial deformity)
(NB. If Choice A is selected the utility is recorded as 0.95)

CHOICE B
Gamble with uncertain outcome
90% chance of perfect health
10% chance of death
(NB. If Choice B is selected the next card is shown)

The principles involved in the SG may be difficult for some respondents to understand and visual aids were used to make the concepts easier. A card deck was constructed using laminated diagrams in which colour-coded pie chart segments represented the probabilities (Figure 2). In order to avoid an anchoring bias, where values were constantly increased or decreased, a ‘ping-pong’ technique was used. This involved alternating values between high and low percentages (10/90, 90/10, 20/80, 80/20, etc.) and reduced the risk of respondents overshooting (i.e. over- or under-estimating the indifference point).

Time Trade-Off
The study assumed the remaining life expectancy to be 50 years, as the average age of the control group respondents was 24.8 years and that of the experimental group 24.0 years. The subjects were offered Choice A (State i), which was 50 years with dentofacial deformity or Choice B, which was a shorter life span, but in perfect health with no deformity (Figure 3). The time in the healthy state (time x) was varied until an indifference point was reached, at which the utility value was recorded. As with the SG, bias was reduced by alternating between long and short periods (for example, 45 years, followed by 5, 40, 10, 35, 15, 30, 20, 25). The more years a respondent was prepared to forego, the lower the utility value (Figure 4).
**Utility Values for Dentofacial Deformity**

**Table 1** Comparison of three methods (including test-retest data).

<table>
<thead>
<tr>
<th>Group</th>
<th>Rating scale</th>
<th>Standard gamble</th>
<th>Time trade-off</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td>Median</td>
<td>Range</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Experimental</td>
<td>57</td>
<td>10-93</td>
<td>59.54 (19.76)</td>
</tr>
<tr>
<td>Control</td>
<td>57</td>
<td>11-88</td>
<td>54.25 (19.03)</td>
</tr>
<tr>
<td><strong>Comparison of the two groups</strong></td>
<td>0.37</td>
<td>0.71</td>
<td>0.27</td>
</tr>
<tr>
<td><strong>Test-retest</strong></td>
<td>Product-moment correlation</td>
<td>0.68*</td>
<td>0.66*</td>
</tr>
<tr>
<td>Measure of agreement</td>
<td>(NS)</td>
<td>(NS)</td>
<td>(NS)</td>
</tr>
</tbody>
</table>

*P < 0.05; **P < 0.01.

**Repeatability of the method**

In order to establish the repeatability of each of the three methods, 20 respondents underwent an identical interview 6 weeks after the first.

**Statistical analysis**

Data were analysed using the SPSS for Windows package (SPSS for Windows, SSPS International BV, Gorinchem, The Netherlands). Both mean and median utility values were calculated to avoid outlying results skewing the data. Comparison of utility values between the two groups used the Mann–Whitney test.

Test-retest repeatability and method agreement were determined using product-moment correlation coefficients, and the measure of agreement method recommended by Bland and Altman (1986).

**Results**

Table 1 shows the median and mean utility values (with standard deviations) for both groups from all three methods. The median value for the RS was 57 (with similar ranges) for both groups. Median values for the SG and the TTO were 0.85 and 0.75, respectively (for both experimental and control groups). However, the range and standard deviations were reduced for the control group. There were no significant differences between the mean values for the two groups for any of the three methods.

**Table 2** Measure of agreement between the RS and TTO with the SG ('gold standard').

<table>
<thead>
<tr>
<th></th>
<th>Product-moment correlation</th>
<th>BSI reproducibility coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG versus RS</td>
<td>0.35**</td>
<td>0.43</td>
</tr>
<tr>
<td>SG versus TTO</td>
<td>0.70**</td>
<td>0.29</td>
</tr>
</tbody>
</table>

**P < 0.01.**

Table 1 also includes the test-retest results. The product-moment correlation was significant for all three methods with the TTO giving the strongest correlation and greatest level of significance (0.84; P < 0.01). For the measure of agreement, the smaller the value, the greater the repeatability. The table illustrates similar levels of repeatability for the SG and TTO (0.09) with a higher value of 0.15 for the RS.

Table 2 illustrates the agreement between the values ascertained from the RS and the TTO when compared with the ‘gold standard’ (the SG). The product-moment correlation was significant for both methods, but the correlation was much lower when comparing the RS versus SG (0.35) than for the TTO versus SG (0.70). The BSI reproducibility coefficient for the SG versus RS was 0.43 and 0.29 for the TTO versus SG.

**Discussion**

One of the major benefits of the utility approach is that it produces a single measure that is amenable to statistical analysis. In addition, it is now an established method based on solid theory.
The question of who should be asked to provide utility values has been the subject of much discussion. Torrance (1986) believed that informed members of the public should be used when public policy decisions were in question, the main factor being that respondents must be accurately informed. The investigation described above achieved this by use of both verbal descriptions and photographs. This study found that there were no significant differences between the utility values for the experimental and control groups. One of the underlying aspects to this study was that it was felt there may be a problem ensuring that members of the public were fully informed when determining utility values for dentofacial deformity and orthognathic treatment. Unlike other medical interventions, such as treatment for arthritis or angina, this is an area with which the majority of lay people may not be familiar. The fact that there were no significant differences between the two groups for any of the three methods of eliciting utility values suggests that members of the public understood the health state descriptions provided and that general public utility values may be considered a satisfactory source of data.

The three methods produced different values, as noted by Mooney (1994). The highest utility value was provided by the SG method and the lowest by the RS. The lower value from the RS reflects findings from previous research (Torrance, 1976). It seems probable that TTO values were lower than those from the SG because, although most respondents were quite willing to consider losing a number of years of their life when they had 50 years to consider, the extent of life/death gamble they were prepared to accept was more limited.

Interpretation of results obtained in utility studies is always difficult. Information is derived primarily by comparing the values with those from other studies. For example, the following utility values were obtained using a conventional TTO method (Torrance, 1987):

1. Some physical and role limitation with occasional pain (0.67).
2. Hospital dialysis (0.56).
3. Anxious and lonely much of the time (0.45).
4. Blind or deaf or dumb (0.39).
5. Hospital confinement (0.33).

Comparing these results with the present study where the mean utility value from the TTO method was 0.67, suggests that dentofacial deformity rates on a par with physical and role limitation with occasional pain, but was not considered as severe a problem as renal dialysis, being anxious or lonely much of the time, or being blind, deaf, or dumb.

A study of oral cancer (Downer et al., 1997) using the conventional SG approach found utility values of:

1. Pre-cancer (assuming full recovery) (0.92).
2. Stage 1 oral cancer (also assuming recovery) (0.88).
3. Stage 2 oral cancer (assuming the cancer is life threatening) (0.68).

Again, when the results of the current study (the SG method gave a utility value of 0.73) are compared with these values, dentofacial deformity is considered a worse health state than either of the lesions associated with recovery, but not as great a problem as when the lesion is life threatening. Obviously, one point which must be borne in mind in this situation is that oral cancer tends to occur in an older age group than the group in this study and, as such, the subjects have to deal with dentofacial problems over much of their life, which may influence the results.

The difficulty of using the SG to look at health states not associated with mortality is in the selection of the ‘worst health state’. However, this is quite a difficult concept for respondents to associate with and it may be that, although the SG is traditionally considered as the gold standard, the TTO may be more appropriate in this type of situation. Subjects in this study did not, however, appear to have difficulty in answering the SG probabilities.

This study used death as the worst health state in the SG and RS for two reasons:

1. It was felt that the oral cavity should be considered in the overall concept of a person’s health.
2. If the anchors used were not 'perfect health' and 'death', the utility values would have extremely limited use because they could not be compared with those for other health states which have used these traditional anchors.

If the SG is accepted as the 'gold standard', validity of the other techniques can be determined by comparison. The product-moment correlation for the SG versus TTO was stronger than for the SG versus RS (0.70 and 0.35, respectively). The BSI reproducibility coefficient provides an indication of the maximum difference likely to occur between two methods or, alternatively, it is the value below which the difference between paired results may be expected to lie with 95 per cent certainty. When the RS and TTO were compared with the gold standard, the BSI reproducibility coefficients were 0.43 and 0.29, respectively. This suggests that utility values derived by different methods are not interchangeable and illustrates the importance of quoting the method used, as well as the utility value obtained. Table 2 also shows the lack of comparability with the RS, which was found in other studies. Torrance (1976) believed that this may be due to the fact that no gamble is involved and suggested that the RS should always be combined with one of the other methods. Torrance (1987) suggested that the RS may be related to the other methods by means of a power curve rather than a linear function. The reason for the lack of agreement between the SG and TTO utility values has been discussed earlier.

The test-retest results in Table 1 show the highest product-moment correlation for the TTO (0.84) with slightly lower values for the other two methods. Similar measures of agreement were found for the SG and TTO (0.09 and 0.09) with a higher value (e.g. lower agreement) for the RS. Few studies report test-retest reliability results although Torrance (1987) quoted values of 0.63 to 0.80 for the TTO at 6 weeks, which is comparable with the present study. Torrance (1976) also reported values of 0.49, 0.53, and 0.62 for the RS, SG and TTO, respectively. However, as the retest was after a 1-year interval lower values would be expected.

The methods used were found to be acceptable to respondents. They appeared to understand the hypothetical situations and argued their utility choices with sound reasoning. There were no refusals to participate in the study and only one set of data for the control group could not be used as the interview was not completed. However, it must be borne in mind that the control group was a convenience sample.

Address for correspondence

S. J. Cunningham
Orthodontic Department
Eastman Dental Institute
256 Gray's Inn Road
London WC1X 8LD, UK

Acknowledgements

The authors would like to express their gratitude to Professor George Torrance of McMaster University, Ontario, and Dr Franco Sassi of the London School of Economics, whose advice in the early stages of this study was invaluable. The authors would also like to thank Drs Aviva Petrie and Ruth Holt for their statistical advice.

References


Mooney G 1994 Key issues in health economics. Harvester Wheatsheaf, New York


Torrance G W 1987 Utility approach to measuring health-related quality of life. Journal of Chronic Disease 40: 593–600