Developing an outcome measure in temporomandibular disorders

Thesis prepared for the degree of
Doctor of Philosophy PhD in Dental Sciences

2007

Mohammad Adel Moufti

Supervisors

Robert W Wassell

Jimmy G Steele

John G Meechan

NEWCASTLE UNIVERSITY LIBRARY
205 36758 8
MED Thesis L8480

In the name of Allah, the Most Gracious, the Most Merciful

Acknowledgment

First and foremost, praise is due to Allah (God) for granting me the opportunity to learn and the ability to complete this thesis.

I would like to acknowledge the generous scholarship I was granted by Damascus University in Syria to prepare this PhD project.

I am also obliged to the patients and individuals who took part in this project, and shared with us their personal feelings, suffering and opinions.

My words cannot thank my supervisors enough for their unlimited support throughout this project. I am so grateful for their time, patience and guidance. I am also thankful for giving me the opportunity to attend various national and international conferences, which had a very positive outcome on the project. Above all, I would like to thank them for believing in me, and continuously encouraging me to move forward.

I am indebted to Jill Smith and Judith Badminton for their friendship and constant help. They both spared no efforts in helping me settle well in England, and in collecting data for this project. No request was considered too small or too much bother.

A sincere thank you to our collaborators: Dr Catherine Exley from the Institute of Health and Society (Newcastle), Dr Finbarr Allen from the Dental School and Hospital (Cork) and Dr Mike John from the School of Dentistry (Leipzig) for the fruitful discussions that enriched this thesis. Their advice was indispensable in the planning and analysis stages of this project. Thanks, too, are extended to the Faculty statistician, Dr Nick Steen and to Dr Simon Kometa from the Computing Department at the University for their advice.

To the library staff at the University, I send a heartfelt thank-you, especially to Rebecca McCready and Erika Gwynnett who accommodated my myriad of requests with smiling

graciousness. I wish also to express my sincere gratitude to the Hospital staff, especially Angela Fenwick and Liz Smith for their help on the clinics.

Last and surely not least, I want to acknowledge my wonderful friends and colleagues. There were periods at my lowest point when someone would purposefully ask, "How is it going?"

The work on this thesis has been an inspiring, often exciting, sometimes challenging, but always interesting experience. It has been made possible by many other people who have supported me. It is impossible to thank every one of them personally. To those whom I did not specifically name, I give my heartfelt thanks for moving me towards my goal.

Table of contents

Acknowledgment I

Table of contents III

List of figures VI

List of tables VII

List of appendices VIII

List of abbreviations IX

Abstract X

Presentations, awards and articles resulting from this work XII

Chapter 1. Introduction to project 1

Chapter 2. Review of the literature 4

2.1. Introduction to the literature review 5

2.2. The Temporomandibular joint and its disorders 6
2.2.1. Temporomandibular joint (TMJ) and masticatory system 6
2.2.2. Temporomandibular disorders (TMD) 7
2.2.3. Temporomandibular disorders – A multidimensional model 10
2.2.4. TMD assessment – the physical dimension 12
2.2.5. TMD assessment – the psychological dimension 22
2.2.6. TMD assessment – the psychosocial dimension 23

2.3. Assessing health in chronic disorders – domains of assessment 28
2.3.1. Rationale for the use of subjective health measurement 29
2.3.2. Applications of subjective health measurement 32
2.3.3. Health components and indicators 33
2.3.4. Core outcome domains for chronic pain research 35

2.4. Assessing health in chronic disorders – methods of assessment 41
2.4.1. Interviews 42
2.4.2. Diaries 48
2.4.3. Questionnaires 53
2.4.4. Scales 67
2.4.5. Psychometric characteristics 74
2.4.6. Other methodological considerations 80

2.5. Assessing health in chronic disorders – timing of treatment outcome assessment 84
2.5.1. The review period 84
2.5.2. The reference period 87
2.5.3. Temporal dynamics 88

2.6. Assessment of oral health-related quality of life (OHrQOL) 89
2.6.1. Oral specific health related quality of life instruments 90
2.6.2. Historical uses of OHrQOL instruments 91
2.6.3. Types of OHrQOL instruments 91
2.6.4. Choosing an OHrQOL measure 94
2.6.5. The Oral Health Impact Profile (OHIP) 97
2.6.6. Short-form versions of OHIP 98
2.6.7. OHIP as a measure of TMD-related quality of life 100

2.7. Summary of the literature review 102

Chapter 3. Aims of project 104

Chapter 4. Study One: Determination of suitable review and reference periods to assess TMD treatment outcomes: Analysis of TMD pain diaries 106

4.1. Introduction *107*

4.2. Aim *109*

4.3. Objectives *109*

4.4. Materials *111*

4.5. Study 1-A: Identifying improvers *112*
4.5.1. Methods 112
4.5.2. Results 119
4.5.3. Discussion 125
4.5.4. Summary 132

4.6. Study 1-B: Examining the general pattern of improvement *133*
4.6.1. Methods 133
4.6.2. Results 138
4.6.3. Discussion 140
4.6.4. Summary 142

4.7. Conclusions of Study One *143*

Chapter 5. Study Two: Determination of suitable review and reference periods to assess TMD treatment outcomes: A qualitative approach – Interviews with TMD patients 144

5.1. Introduction *145*

5.2. Aim *147*

5.3. Objectives *147*

5.4. Methods *148*
5.4.1. Subjects 148
5.4.2. Preparatory stage 149
5.4.3. The interviews 149
5.4.4. Data analysis 150

5.5. Results *152*
5.5.1. Symptoms of TMD and their improvement 152
5.5.2. The impact of TMD on quality of life and its improvement 158
5.5.3. Time scale of improvement 162
5.5.4. Perception of care 164

5.6. Discussion *168*
5.6.1. Symptoms of TMD and their improvement 169
5.6.2. The impact of TMD on quality of life and its improvement 171
5.6.3. Time scale of improvement 172
5.6.4. Perception of care 173
5.6.5. Strengths and limitations 174

5.7. Conclusions *175*

5.8. Summary *176*

Chapter 6. Study Three: Establishing a quality of life measure of TMD treatment outcomes 177

6.1. Introduction *178*

6.2. Aim *180*

6.3. Objectives *180*

6.4. Methods *182*
6.4.1. Data collected 182

6.4.2. Study subjects 185
6.4.3. Preparatory stage and pilot work 189
6.4.4. Data entry and cleaning 189
6.4.5. Data analysis: comparing TMD patient and control groups 191
6.4.6. Identifying candidate items (OHIP/TMD-I) 193
6.4.7. Exploring validity and reliability of OHIP and OHIP/TMD-I 194

6.5. Results *199*
6.5.1. Missing data 199
6.5.2. Differences between TMD patient and control groups 200
6.5.3. Candidate items (OHIP/TMD-I) 210
6.5.4. Validity and reliability of OHIP and OHIP/TMD-I 217

6.6. Discussion *221*
6.6.1. Methodological issues 221
6.6.2. Differences between TMD patient and control groups 224
6.6.3. Candidate items (OHIP/TMD-I) 228
6.6.4. Validity and reliability of OHIP and OHIP/TMD-I 232
6.6.5. Strengths and limitations 236

6.7. Conclusions *238*

6.8. Summary *239*

Chapter 7. Conclusion of project and further work **240**

7.1. Conclusions *241*

7.2. Recommendations for further investigation *243*
7.2.1. Study One 243
7.2.2. Study Two 244
7.2.3. Study Three 244

Bibliography **245**

Appendices **263**

List of figures

Figure 2.1. Anatomy of the temporomandibular joint 6
Figure 2.2. A sketch of the most common signs and symptoms in TMD 8
Figure 2.3. Summary of the bio-psychological concept of TMD 11
Figure 2.4. Locker's conceptual model for measuring oral health 90
Figure 4.1. A time scale representing review and reference periods 107
Figure 4.2. Outline of *Study One* 110
Figure 4.3. Different types of pain/time change with treatment 113
Figure 4.4. Detailed outline for the methods used to discriminate improvers from non-improvers 114
Figure 4.5. A plot showing the time periods involved in the mathematical assessment 116
Figure 4.6. Number of days analgesics used during treatment 124
Figure 4.7. Pattern of daily pain changes for an average improver 134
Figure 4.8. Regression analyses of pain scores for an average improver 135
Figure 4.9. Pattern of pain changes for an average improver on a weekly and fortnightly basis 136
Figure 4.10. Weekly means and standard errors of means for improvers 136
Figure 4.11. Concept of the review and reference periods 137
Figure 4.12. Pain reduction with treatment for an average improving patient 138
Figure 4.13. Determination of the reference period 139
Figure 6.1. Outline of *Study Three* 181
Figure 6.2. Distribution of patients and controls in the different age and gender groups 186
Figure 6.3. *Item-Impact* scores for individual items for patients and controls 207
Figure 6.4. Item differences between patients and controls in *Mean*, *Median* and *Prevalence* 212
Figure 6.5. Global rating of oral health 218

List of tables

Table 2.1. Assessment criteria for TMD diagnostic systems 14
Table 2.2. Research diagnostic criteria for TMD (RDC/TMD) 17
Table 2.3. Jaw disability checklist 21
Table 2.4. An example of WHO's classification of functional disability 38
Table 2.5. Summary of strengths and weaknesses of qualitative research 43
Table 2.6. Advantages and disadvantages of the main modes of questionnaire administration 54
Table 2.7. Major considerations in index construction and validation 61
Table 2.8. Values of Pearson and Spearman's correlation and their interpretation 75
Table 2.9. Values of Cohen's Kappa and their interpretation 78
Table 2.10. "Generic" oral health related quality of life instruments 93
Table 4.1. Diagnoses of patients participating in the original trial in *Study One* 112
Table 4.2. Comparison between results from the three visual assessments 120
Table 4.3. Agreement between examiners (Kappa test) for the second and third assessments 120
Table 4.4. Cross-tabulation of TMD treatment group with visually assessed treatment outcome 120
Table 4.5. Corroboration of the visual and mathematical assessments to derive definite improvers 122
Table 4.6. Sensitivity & specificity of the mathematical assessment against the third visual assessment 123
Table 5.1. Characteristics of the patient sample in *Study Two* 148
Table 5.2. Expected and emerging themes obtained from the interviews 152
Table 5.3. Participants' own accounts of jaw pain and headaches 154
Table 5.4. Participants' own accounts of non painful joint-related symptoms 155
Table 5.5. Participants' own accounts of muscle-related symptoms 156
Table 6.1. Distribution of patients within different diagnostic groups 186
Table 6.2. Mann-Whitney test and significance of score differences between patients and controls 201
Table 6.3. Differences in item *Mean* scores between patients and controls 203
Table 6.4. Differences in item *Median* scores between patients and controls 204
Table 6.5. Differences in item *Prevalence* (%FOVO) between patients and controls 206
Table 6.6. Differences in *Item-Impact* values between patients and controls 208
Table 6.7. Average scores for OHIP domains and total OHIP in patients and controls 209
Table 6.8. Percentage reporting FOVO impact to one or more items across domains and groups 210
Table 6.9. Items with no statistically significant difference between patients and controls 211
Table 6.10. Items showing the largest difference between patients and controls 213
Table 6.11. Domains of the items with largest difference between patients and controls 214
Table 6.12. Additional items and their ranking in the four analyses 214
Table 6.13. Summary scores and domains for OHIP/TMD-I 215
Table 6.14. OHIP/TMD-I 216
Table 6.15. MPI summary measures for patients 218
Table 6.16. Correlation between MPI, full OHIP and OHIP/TMD-I 219
Table 6.17. Correlation between full OHIP, OHIP/TMD-I, pain VASs and global rating of oral health 219
Table 6.18. Internal consistency for OHIP with low correlation 220
Table 6.19. Comparison between candidate items and items chosen by Murray *et al.* (1996) 231

List of appendices

Appendix 1 TMD diagnostic criteria and examination forms 264
Appendix 2 Formula for calculating the area under the curve in *Study One* 271
Appendix 3 Individual plots of pain/time response to treatment in *Study One* 273
Appendix 4 Participant information sheet and consent form used in *Study Two* 279
Appendix 5 Topic guide used for the interviews in *Study Two* 283
Appendix 6 Matrices used to map and summarise themes emerging from interviews in *Study Two* 285
Appendix 7 The Oral Health Impact Profile - adapted version 311
Appendix 8 The Multidimensional Pain Inventory 318
Appendix 9 Participant information sheet and consent form used in *Study Three* 323
Appendix 10 Form used to check inclusion criteria for controls in *Study Three* 329

List of abbreviations

HrQOL	Health-related Quality of Life
MPI	Multidimensional Pain Inventory
NDH	Newcastle Dental Hospital
OHIP	Oral Health Impact Profile
OHrQOL	Oral Health-related Quality of Life
QOL	Quality of Life
RDC	Research Diagnostic Criteria
SPSS	Statistical Package for the Social Sciences
TMD	Temporomandibular Disorders
VAS	Visual Analogue Scale
WHO	World Health Organization

Abstract

Assessment of treatment outcome in clinical trials of temporomandibular disorders (TMD) lacks validated patient-based measures of the condition and its improvement. Such measures are questionnaire-based. The questions need to be chosen and framed appropriately. In line with the guidelines of the National Institute for Health and Clinical Excellence in the U.K., this project aimed to start the development of a patient-derived quality of life instrument to measure TMD treatment outcome. Three studies were conducted for this purpose.

Study One aimed to establish the optimal time to assess treatment outcome for TMD (the *Review Period*), and the best *Reference Period* to determine the time-frame of questions. Data consisted of daily diaries of pain intensity from 72 patients who had participated in a clinical trial of conservative TMD treatment. The study involved two stages. The first aimed to identify patients responding to treatment, so-called *"Improvers"* and to exclude the *"Non-Improvers"* whose data might dilute the results. This was accomplished in two ways: firstly by visually assessing individual plots of pain versus time; secondly by mathematically calculating the reduction of pain scores during treatment. Criteria for improvement were set and tested for both methods. Definite improvers were selected if they met criteria of both assessments. In the second stage, only improvers' data were examined to analyse the general trend of improvement and establish the review and reference periods. *Study One* determined 15 weeks as the best review period and four weeks as the best reference period when assessing a commonly used TMD treatment.

Study Two was intended to triangulate with *Study One* and had the same aim. *Study Two* explored, using a qualitative approach, patients' perception of improvement of TMD symptoms and rates of their recovery. Ten TMD patients showing improvement to conservative treatment were interviewed. Data were collected using a pre-designed topic guide and analysed using the *Framework* approach. The interviewees consistently reported

that pain was the symptom with the most impact on their quality of life, and the one to which improvement was linked the most. Other TMD symptoms were also important, but were less linked to patients' suffering and to perceived improvement. Different symptoms followed various rates of recovery. It was not possible to determine a definite time when all symptoms are considered improved. This study confirmed that the approach used in *Study One*, i.e. assessing TMD recovery based on measurement of pain intensity, was reasonable. It also confirmed that sufficient time (in a frame of some months) is needed before outcome for TMD conservative treatment can be assessed meaningfully. In addition, this study identified important new themes related to patient's journey throughout illness and their perception of received care.

Study Three was the core study in this project. It aimed to establish a quality of life measure of TMD treatment outcome using the Oral Health Impact Profile (OHIP) as a starting point, using the reference period identified in studies one and two. A short form of OHIP was derived by identifying OHIP items with the largest impact on TMD patients' quality of life. A case-control design was used and 110 patients (PG) undergoing a variety of conservative treatments and matching controls (CG) were included. Candidate questions (items) for the intended short form (OHIP/TMD-I) were selected based on three criteria: 1- showing statistical differences between PG and CG; 2- showing largest score difference (PG:CG) as measured by four analyses: *Mean, Median, Prevalence* and *Item-Impact*; 3- representing all domains of OHIP's theoretical framework. This was followed by testing aspects of validity and reliability of the short form and comparing them to those of the full OHIP. Twenty items were identified. Their psychometric properties were comparable or better than those of the mother instrument.

Future work is needed to analyse the responsiveness to change with treatment of the short form, to investigate whether items outwith OHIP are needed to furnish the intended TMD outcome instrument, and to test the psychometric properties of the final instrument in a new sample.

Presentations, awards and articles resulting from this work

Presentations

Moufti MA, Steele JG, Meechan JG, Wassell RW (2003). Assessment of treatment outcomes for the TMD: review & reference periods. The British Society for Restorative Dentistry (BSRD). London, U.K.

Moufti MA, Steele JG, Meechan JG, Exley C, Wassell RW (2005). Quality of life and temporomandibular disorders: what do we know? The British Society for Dental Research (BSDR). Dundee, U.K.

Moufti MA, Steele JG, Meechan JG, Wassell RW (2005). Patterns of chronic pain recovery: assessment and classification. 11th World Congress on Pain, International Association for the Study of Pain (IASP). Sydney, Australia.

Moufti MA, Steele JG, Meechan JG, Wassell RW (2005). Treatment of temporomandibular disorders: when should the outcome be assessed? 4th International Conference on Orofacial Pain and TMD. Sydney, Australia.

Moufti MA, Steele JG, Meechan JG, Wassell RW (2006). Developing a measurement of treatment outcomes in temporomandibular disorders. 3rd Meeting of the Pan European Federation for Dental Research (PEF). Dublin, Ireland.

Steele JG, Moufti MA, Durham JA, Meechan JG, Wassell RW (2006). Does TMD affect quality of life in unexpected ways? 3rd Meeting of the Pan European Federation for Dental Research (PEF). Dublin, Ireland.

Moufti MA, Wassell RW, Meechan JG, Allen PF, John MT, Steele JG (2007). OHIP as a measure for TMD-related quality of life. 54th Annual Conference of the British Society for the Study of Prosthetic Dentistry. Newcastle, U.K.

Moufti MA, John MT, Allen PF, Meechan JG, Wassell RW, Steele JG (2007). Sensitivity of OHIP as an outcome measure for temporomandibular disorders. 85th Meeting of the International Association for Dental Research (IADR). New Orleans, USA.

Durham JA, Exley C, Wassell RW, Moufti MA, Steele JG (2007). A comparison of patient and professional perspectives on temporomandibular disorders. The British Society for Dental Research (BSDR). Newcastle, U.K.

Awards

1. *Prize for the Best Postgraduate Research Presentation in the annual meeting for the British Society for Restorative Dentistry (BSRD). London, U.K.*
   Moufti MA, Steele JG, Meechan JG, Wassell RW (2003). Assessment of treatment outcomes for the TMD: review & reference periods. The British Society for Restorative Dentistry (BSRD). London, U.K.

2. *Conference prize for the Best Postgraduate Research Presentation in the TMD management group, 4th International Conference on Orofacial Pain and TMD. Sydney, Australia.*
   Moufti MA, Steele JG, Meechan JG, Wassell RW (2005). Treatment of temporomandibular disorders: when should the outcome be assessed? 4th International Conference on Orofacial Pain and TMD. Sydney, Australia.

3. *The "Behavioral Sciences & Health Services Research Outstanding Student Abstract Award", 85th General Session & Exhibition of the International Association for Dental Research (IADR), New Orleans, USA.*
   Moufti MA, John MT, Allen PF, Meechan JG, Wassell RW, Steele JG (2007). Sensitivity of OHIP as an outcome measure for temporomandibular disorders. 85th Meeting of the IADR, New Orleans, USA.

Articles

Wassell RW, Moufti MA, Meechan JG, Steele JG (2007). A method of defining "improvers" in chronic pain studies. J Orofacial Pain: Submitted.

Moufti MA, Meechan JG, Steele JG, Wassell RW (2007). Treatment of temporomandibular disorders: when should the outcome be assessed? J Orofacial Pain: In Progress.

Chapter 1.
Introduction to project

Over the past decades, emphasis has been increasingly given in health research to use a comprehensive biopsychosocial model of "illness" in lieu of the old biomedical model of "disease" that has failed to account for the various psychological and psychosocial variables, essential elements in the aetiology and recovery of illness (Dworkin and Massoth, 1994). Health care regulatory bodies recommend that patients' quality of life (QOL) should be considered as a principal element of outcome assessment (National Institute for Health and Clinical Excellence, 2006). They stress the fact that what matters is how patients feel and not how professionals think they feel. This becomes especially essential when dealing with chronic disorders, where patients' experience and suffering extend over long periods.

Temporomandibular Disorders (TMD) are recognised as chronic disorders affecting the temporomandibular joints, muscles of mastication or both. These, like other chronic pain conditions affect patients' quality of life, and should be assessed using a multi-axial biopsychosocial approach. There is a lack, however, of outcome measures that incorporate this model in research on TMD clinical trials.

This project aimed to start the development of a patient-based quality of life measure of TMD treatment outcome. It involved three stages. The first and second stages shared the same aim, which was to establish the optimal time to conduct the outcome assessment (*Review Period*), and the best time frame for patients to retrospectively assess impact of treatment on quality of life (*Reference Period*). Using quantitative methods in *Study One* and a qualitative approach in *Study Two*, data from patients undergoing conservative TMD treatments were analysed to examine rates of recovery and to consequently determine the above review and reference periods.

The third stage of the project sought to develop a quality of life measure of TMD treatment outcome, based on a well established measurement of oral health-related quality of life (OHrQOL), the Oral Health Impact Profile (OHIP) (Slade and Spencer, 1994). OHIP has

been successfully used to measure the impact of TMD and orofacial pain on quality of life (Murray *et al.*, 1996; John *et al.*, 2002; Larsson *et al.*, 2004; Segù *et al.*, 2005). However, many of OHIP's 49 items are apparently unrelated to TMD and would compromise its utility as a measure of TMD treatment outcome. Empirical attempts have been made to produce a TMD-related version by removing apparently items not related to TMD and by adding items that are considered relevant (Murray *et al.*, 1996). That work, however, was not based on a strong scientific basis. Validated methods are available to distil long instruments, and have been used to derive condition-specific versions of OHIP, such as that designed for edentulous patients (OHIP-EDENT) (Allen and Locker, 2002).

The main thrust of this present work was to derive a TMD-specific version of OHIP using a rational process, taking into consideration the time-frame identified in *Study One* and *Study Two*. Having developed a prototype of the instrument, some aspects of its validity and reliability have been explored.

The present project completed the principal stage of the development of a TMD quality of life outcome instrument. Further work is needed to examine sensitivity to change with treatment of the proposed short form, to explore whether items additional to OHIP are needed and to establish the psychometric properties of the final product.

Chapter 2.
Review of the literature

2.1. Introduction to the literature review

This review consists of five sections. It starts with a section on the definition of the temporomandibular disorders (TMD) and their position within other chronic illnesses. The section introduces the concept of assessing these conditions taking into consideration psychological and psychosocial factors as well as the physical element of the illness. Examples of instruments used for the assessment of TMD are presented with a report on the lack of specific quality of life related measures.

Section two looks at TMD and other chronic disorders from a wider angle. It considers these conditions under the general umbrella of health, and stresses the need to assess these conditions within the multidimensional concept of health and illness. This section expands more on the importance of patient-based assessment of own health and on the concept of quality of life. It outlines the different dimensions under which treatment outcome should be evaluated, especially in chronic pain conditions.

Section three illustrates the different patient-based methods of health assessment. Both qualitative and quantitative methods are considered, including their advantages, disadvantages, proper applications and psychometric properties.

Timing of outcome assessment is an essential element in the success of such assessment. Section four examines the importance of two time periods in relation to outcome assessment: the optimal time to conduct this assessment (the *Review Period*) and the time-frame used in the assessment (the *Reference Period*). This section also reports the lack of validated review and reference periods for TMD.

The last section reviews the literature for available measures of oral health-related quality of life (OHrQOL), with a brief examination of their types and applications. This is followed by a detailed review of the Oral Health Impact Profile (OHIP), the instrument used in this project.

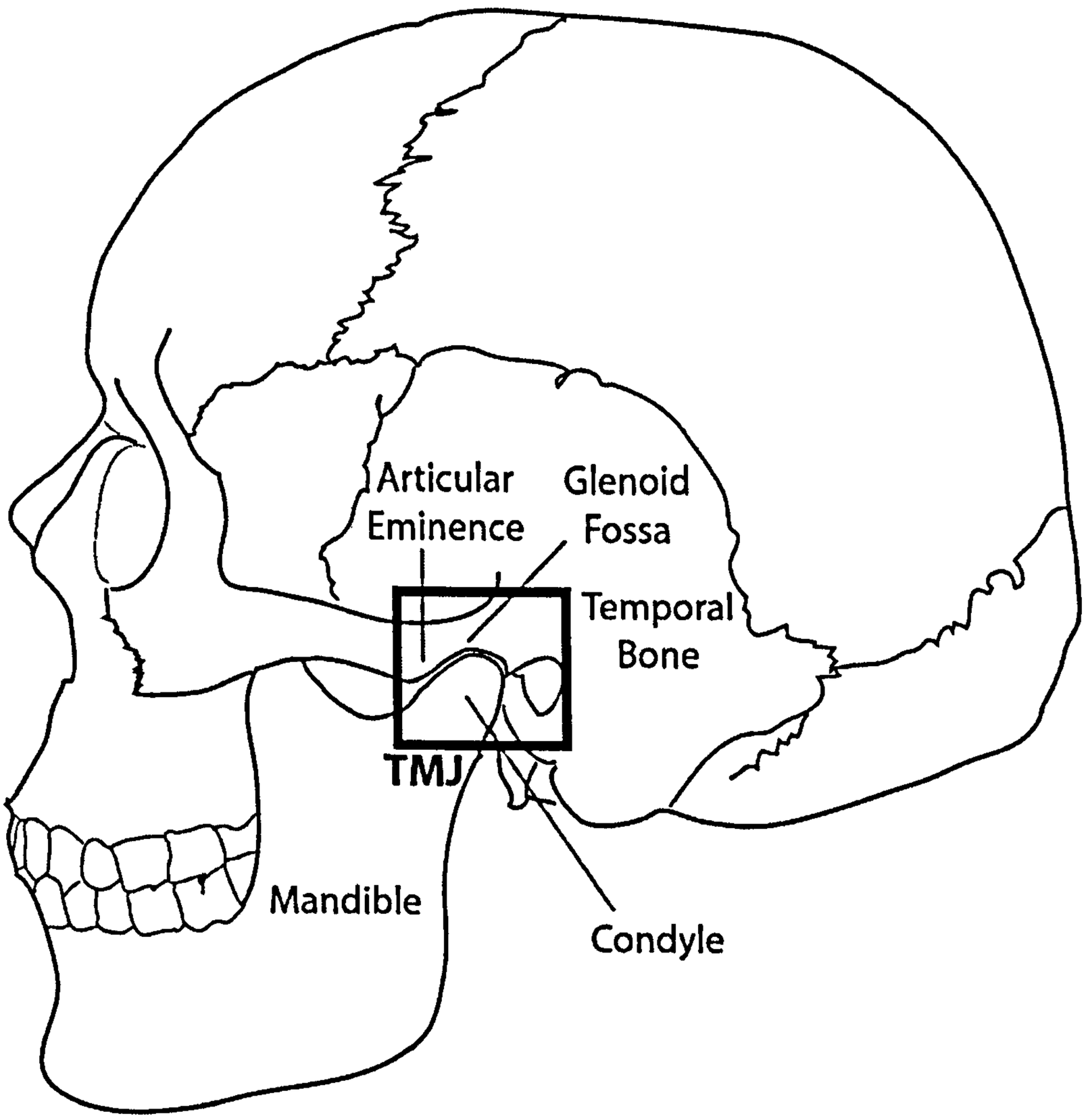
This review is concluded with a summary of the identified key points.

2.2. The Temporomandibular joint and its disorders

2.2.1. Temporomandibular joint (TMJ) and masticatory system

The TMJs provide the articulation of the mandible with the cranium (Figure 2.1), more specifically between the mandible condyle and the squamous part of the temporal bone, which are separated by a thin fibrous disc (Okeson, 2003). Movement of the mandible around the right and left joints is provided by means of a complex system of muscles and nerves. In addition, both the mandible and the disc are supported with ligaments to limit their movements. Teeth also play an important role in determining the path of the joint movements. This complex system of bones, discs, teeth, muscles, ligaments and nerves is referred to as the masticatory system.

Figure 2.1. Anatomy of the temporomandibular joint



The anatomical complexity of the joint provides compound functions. It enables the mandible to perform a large number of "food-free" movements such as opening, closing and excursion; and food-related (or masticatory) functions including incision and mastication. This is all regulated by the coordinated function of the muscles surrounding the mandible. Two types of movements are facilitated by the double-faced articulation of the disc, namely rotation of the condyle against the disc, and translation of the disc against the temporal bone. The TMJ is therefore classified as a hinge joint with a movable socket.

Normal structure and function of this inter-correlated system are maintained through the harmoniously co-ordinated function of its parts. However, with this high level of complexity both in anatomy and function, it is not uncommon for "incidents" to occur, where some parts become affected by external factors. Not only may these components start to malfunction, but other parts of the system may become affected indirectly, often leading to a complex manifestation of signs and symptoms referred to as Temporomandibular Disorders (TMD).

2.2.2. Temporomandibular disorders (TMD)

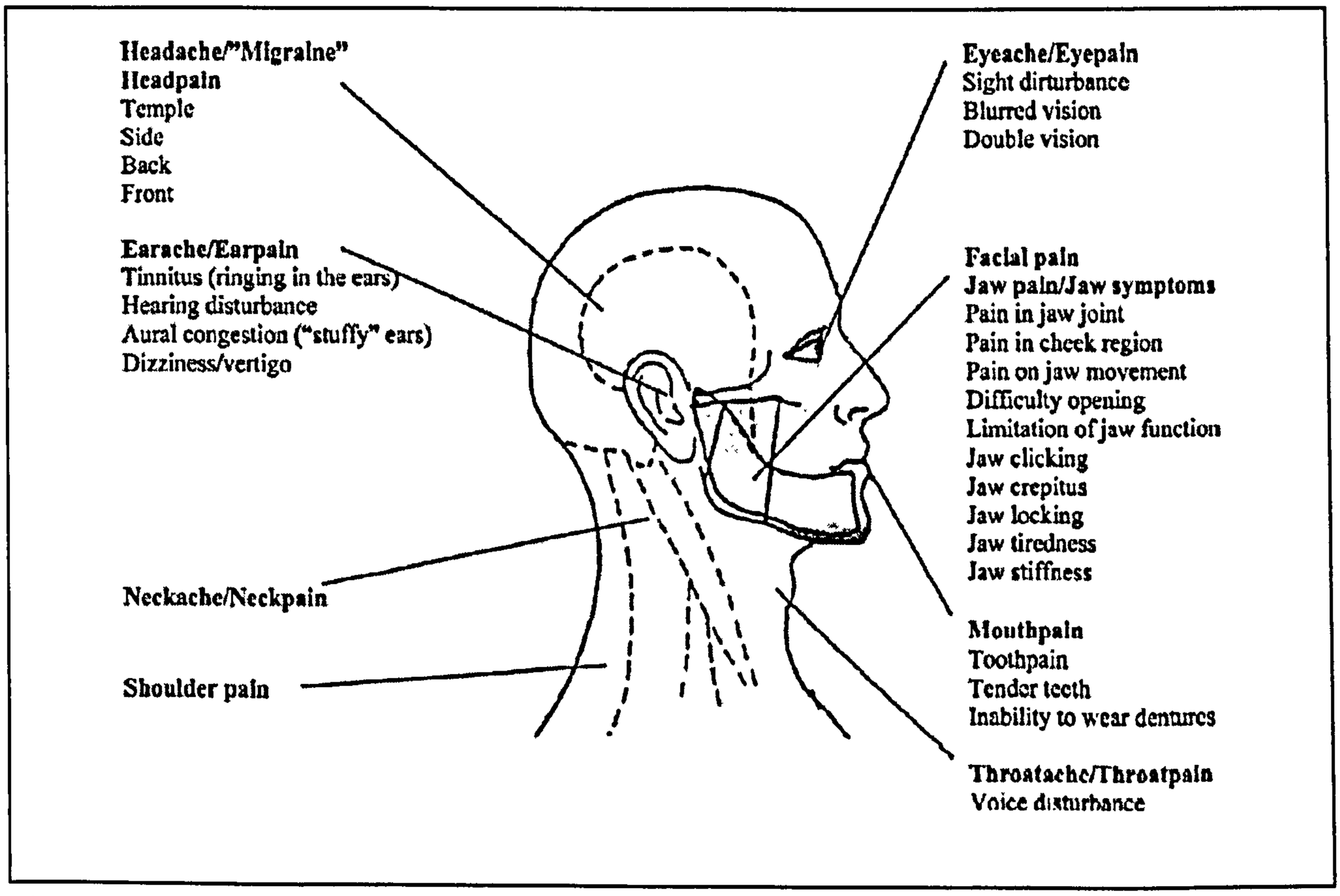
The first article on TMD appeared in 1934, when James Costen, an otolaryngologist, reported "A Syndrome of Ear and Sinus Symptoms Dependent Upon Disturbed Function of the Temporomandibular Joint" (Costen, 1934). Since then a vast amount of research has been carried out on the TMJ and disorders of the masticatory system. But, because of the various conditions affecting these structures, workers, having come from different clinical and scientific backgrounds have failed to reach a consensus on nomenclature and definition. In this thesis I have chosen that of the American Academy of Orofacial Pain (McNeill *et al.*, 1993):

> *"Temporomandibular disorders (TMD) is a collective term embracing a number of clinical problems that involve the masticatory musculature, the temporomandibular joint (TMJ) and associated structures, or both".*

TMD presents clinically with different symptoms (Figure 2.2) including pain, clicking or grinding sounds in the joint; dysfunction and limitation in mouth opening and other movements. The pain is sometimes localised in the joint and/or the masticatory muscles, and at other times radiates to neighbouring areas to produce earache, headache, toothache and pain in the face and neck.

TMD often co-exists with other craniofacial and orofacial pain disorders. It is considered a sub-classification of musculoskeletal disorders, including arthritis and fibromyalgia (Feine and Lund, 1997; Aaron *et al.*, 2000)

Figure 2.2. A sketch of the most common signs and symptoms in TMD*



*** From (Suvinen *et al.*, 2005)**

Epidemiological studies on TMD are contradictory about its prevalence because of disagreement on the descriptive terminology, assessment and data collection methodology (Dworkin *et al.*, 1990; LeResche, 1997b). A review of epidemiologic studies of TMD has estimated that 28-75% of the general population exhibit at least one sign e.g. joint noise or

tenderness to palpation and 33% have at least one symptom (Rugh and Solberg, 1985). These authors, however, reported that only 5% of the general population are in need for treatment. More recent studies have explored the prevalence of individual TMD signs and symptoms as opposed to that of TMD as a distinctive condition. Various signs and symptoms were found to differ in prevalence both amongst the patient and the general population (Dworkin *et al.*, 1990; Goulet *et al.*, 1995; LeResche, 1997b). For example, amongst those seeking treatment for TMD, pain was found to be the most prevalent symptom (97%) compared to 43% presenting joint clicking and 20% exhibiting limited mouth opening (Dworkin *et al.*, 1990). These signs were less common, but still exist in subjects without a clinical diagnosis of TMD. For example, the same study reported pain prevalence in the TM region (TMJ and relevant muscles) in 12% of the general population and joint clicking in 25%. Other joint sounds, i.e. crepitus and "grating" were found far less common both in the patient and general population (Dworkin *et al.*, 1990). A more recent study on the prevalence of TMD signs and symptoms found that jaw pain, restricted jaw opening, and jaw clicking are the most common presenting symptoms amongst TMD patients, but reported different prevalence (Pow *et al.*, 2001) (see below).

While authors agree that signs associated with TMD occur quite commonly in the general population, they note that only 5-7% of the population seek evaluation and care (Rugh and Solberg, 1985; Schiffman *et al.*, 1990; McNeill *et al.*, 1993; Goulet *et al.*, 1995). These figures seem to be inconsistent across different populations. For example, the data from the USA reported above are higher than recent findings in a Hong Kong Chinese population. In the latter only 1 % reported TMD-related jaw pain of moderate or severe intensity and frequent occurrence, and only 0.6% had sought treatment for jaw pain, impaired jaw opening, or joint clicking that occurred often in the previous year (Pow *et al.*, 2001).

Prevalence of signs and symptoms in men and women varies between the general and patient populations. While the ratio of women to men is reported to be 1:1 in the non-patient population (Helkimo, 1974b), studies on subjects seeking treatment showed that this ratio is between 3:1 and 9:1 (McNeill *et al.*, 1993; LeResche, 1997b; Johansson *et al.*, 2003). In relation to age, TMD sufferers are reported to be primarily young and middle-aged adults (McNeill *et al.*, 1993; LeResche, 1997b).

Although TMD is usually a self-limiting condition, a significant number of patients develop a chronic pain syndrome (Dworkin *et al.*, 1990; McNeill *et al.*, 1993). This is defined as pain persistent for more than six months with associated behavioural and psychosocial factors (McNeill *et al.*, 1993).

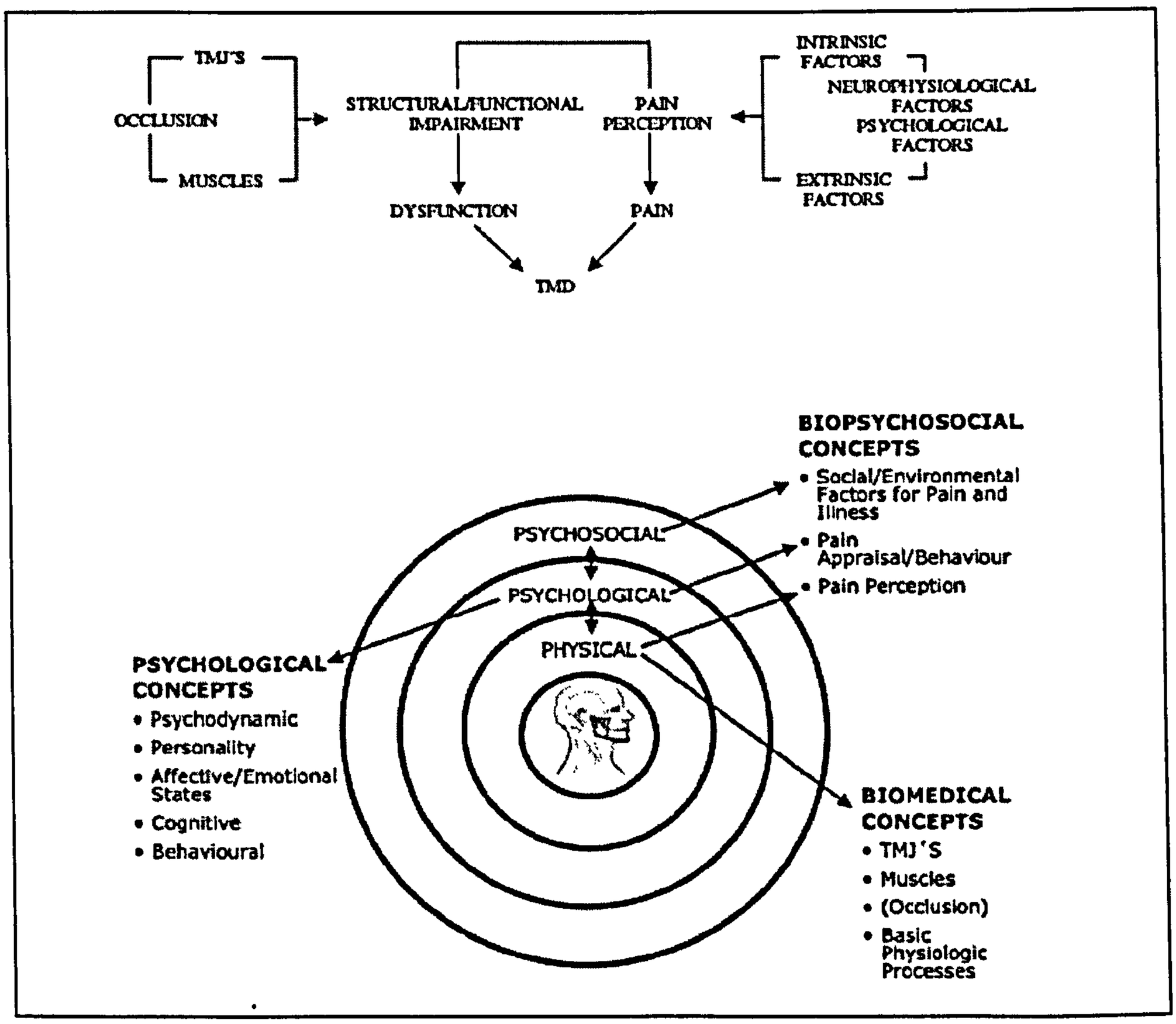
2.2.3. Temporomandibular disorders – A multidimensional model

Chronic pain patients respond to pain differently from acute pain patients (McNeill *et al.*, 1993). Ongoing (chronic) pain perception is exacerbated by other psychological factors including social situations, emotions (anger, fear) and attitudes (Kerns *et al.*, 1985; Dworkin and LeResche, 1992; Turk, 1997; Crombez *et al.*, 1999; Holroyd *et al.*, 1999; Dworkin *et al.*, 2002a; Turk *et al.*, 2003; Suvinen *et al.*, 2005). The psychosocial dimension of chronic pain is well documented in the literature, and is covered in more detail in section 2.3.4.

Accumulated evidence indicates that TMD shares the major characteristics of other chronic pain conditions such as headache and back pain (Dworkin and Massoth, 1994). Increasing attention has been given to the importance of studying TMD using the same biopsychosocial model used for common pain conditions (Dworkin and Massoth, 1994; Suvinen *et al.*, 2005). The model encompasses three dimensions: the physical, psychological and psychosocial element (Figure 2.3). The need for this broad model to measure health and treatment outcomes is explored in more detail in section 2.3. The remainder of section 2.2,

however, reviews the various methods for the assessment of TMD patients within this model. Available instruments for assessment of the condition and treatment outcomes are discussed under the respective physical, psychological and psychosocial dimensions.

Figure 2.3. Summary of the bio-psychological concept of TMD*



*** From (Suvinen *et al.*, 2005)**

It should be noted, however, that while this model is extensive, it does not take into consideration the recent recognition of the role of genetics as an internal factor explaining the variations among patients in pain perception (Diatchenko *et al.*, 2005).

The following sections review the literature on the assessment of TMD. While measurement of pain in general and in relation to TMD is reported elsewhere (sections 2.3.4 and 2.4.4.4), pain-related impact on patients' lives is reported here.

2.2.4. TMD assessment – the physical dimension

This section concerns the first dimension of the ternary model previously described. It reviews instruments used to diagnose TMD and monitor outcomes in terms of physical signs, functional ability and functional limitation of the masticatory system.

2.2.4.1. TMD diagnostic systems

Diagnostic systems provide either *Case Definitions* (i.e. classification and diagnostic criteria) or *Examination Protocols*, or both. Numerous TMD diagnostic systems have been developed since Bell first published his taxonomy in 1970 (Bell, 1970):

1. Bell's classification of TMD (Bell, 1970)
2. Farrar's taxonomy for TMJ Dysfunction Syndrome (Farrar, 1972)
3. Helkimo's Dysfunction Indices (Helkimo, 1974a)
4. Block's classification of craniofacial-cervical pain (Block, 1980) cited by Dworkin and LeResche (1992)
5. The American Academy of Craniomandibular Disorders' classification (McNeill *et al.*, 1980; McNeill, 1983)
6. Eversole and Machado's classification of "TMJ musculoligamentous derangements" (Eversole and Machado, 1985)
7. The Craniomandibular Index, CMI (Fricton and Schiffman, 1986)
8. The TMJ Scale (Lundeen *et al.*, 1986; Levitt and McKinney, 1994)
9. The American Academy of Head, Neck, Facial Pain, and TMJ Orthopaedics (AAHNFP& TMJO)'s classification (Talley *et al.*, 1990)
10. The Clinical Diagnostic Criteria, CDC (Truelove *et al.*, 1992)
11. The International Headache Society Classification (McNeill *et al.*, 1993)
12. The Research Diagnostic Criteria, RDC (Dworkin and LeResche, 1992)
13. The Temporomandibular Index, TMI (Pehling *et al.*, 2002).

Different attempts have been made to review and evaluate these systems, e.g. (Dworkin and LeResche, 1992; Clark *et al.*, 1993; Okeson, 1997). The most comprehensive of these reviews is that conducted by the group who later established the RDC (Dworkin and LeResche, 1992). The scope of that review was to determine the performance of existing systems at that time with the final aim to establish a scientific diagnostic system that is valid for research purposes. The authors firstly decided the features that a scientific TMD classification system should have. These criteria consisted of methodological considerations and clinical applications (Table 2.1 below) and were used to examine existing diagnostic systems. In addition, the team took into consideration the comprehensiveness of patient assessment, predictive value of making a diagnostic assignment, and implications for choice of treatment.

Nine diagnostic systems were included in the RDC review. These were systems that had a specific classification of TMD as their purpose. Systems that were not oriented toward TMD particularly (i.e. toward diseases or pain in general) or not formulated to collect the information which met RDC requirements (i.e. standard of evaluation or TMD diagnostic criteria) were excluded.

Table 2.1. Assessment criteria for TMD diagnostic systems *

Criterion	Description	Rating
Methodological Considerations		
Sample method [1]	Study design for testing diagnostic criteria	Case-series vs case control Cross-sectional vs longitudinal Prospective vs retrospective
Sample type [2]	Source of subjects used in testing diagnostic criteria	Population Clinical Unknown
Research suitability	Whether criteria are stated in measurable terms	Yes No
Inter-rater reliability (IRR) method [3]	IRR for evaluation methods, according to whether data are provided by the proponents of the system and whether all evaluation methods have IRR support	Internal-full Internal-partial External-full External-partial
Specificity [4]	Whether diagnostic criteria of a system detect "disease" in a non-patient population	Acceptable Unacceptable Unknown
IRR diagnosis [5]	IRR for whether different judges would make the same diagnostic assignment	Acceptable Unacceptable Unknown
Clinical Considerations		
Biological	Whether the system is compatible with current anatomical, behavioural, and physiological knowledge	Strong Moderate Minimal
Exhaustive	Whether the system can classify all known clinical presentations	All Major Minor
Multiple diagnoses	Whether multiple diagnoses are allowed	Yes No Unknown
Decision making [6]	Whether system is organized to facilitate decision making	Good Poor

*** From (Dworkin and LeResche, 1992)**

[1] If assessed as "None": study method is not identified
[2] The term "None" is used where no particular sample is used in the study
[3] Acceptable IRR are granted when a Kappa statistic was 6.0 or above. "Unknown IRR" is given to measures which have no IRR or are insufficiently operationalized
[4] The value of 0.75 determined whether or not the specificity is acceptable, and the *"Unknown"* term is used to describe lack of data
[5] A Kappa value of 0.6 is used to determine the reliability. The term *"Unknown"* is given when data were not available
[6] Systems must meet the following to facilitate decision making:

1. To be organized in a decision-tree concept
2. Essential clinical characteristics required for assessing a diagnosis are clear
3. A distinct diagnosis can be obtained through the criteria of each diagnostic category.

Based on the above-mentioned assessment criteria the RDC team found that:

1. All systems, except Truelove's, contain statements which are not essential for the final diagnosis, which lead to a difficulty in the use of these taxonomies
2. There were differences among diagnostic systems in classifying each TMD entity in the context of each system. Furthermore, there were contradictions among systems in determining specific TMD entities. RDC gives as an example behavioural phenomena such as "parafunctioning"
3. RDC defined four problems regarding the terminology of TMD entities. These were:
   a. The lack of anatomical or physiological support for some hypothetical terms used in some systems. Examples for that were *"Contracture"* and *"Myospasm"*
   b. The use of statements that have causal implications to describe some terms (e.g. muscle splinting)
   c. The internal inconsistency between how a disorder is defined and what the stated essential criteria are
   d. The use of non-descriptive terms for some disorders (e.g. *Arthritis*) in a number of the diagnostic systems
4. All diagnostic systems lacked the psychological assessment of patients. Nevertheless, the RDC recommended the need for further efforts to determine the essential variables in evaluating the psychological "dimension" of TMD patients, to classify these variables, and to agree on the best instrument for assessing this dimension
5. The RDC team believed that aetiology, prognostic statements and implications for treatment are critical issues in assessing nomenclatural systems. Yet, since there is lack of knowledge in these domains, diagnostic systems cannot be judged properly.

In this project, I derived the physical diagnostic criteria from the RDC, and the examination form from the Temporomandibular Index TMI (Pehling *et al.*, 2002). Details about these two systems are presented.

2.2.4.2. The Research Diagnostic Criteria (RDC)

RDC or RDC/TMD (Dworkin and LeResche, 1992) is one of the best known TMD diagnostic systems internationally (Yap, 2002). Unlike other diagnostic systems, both diagnostic components, i.e. case definitions and examination procedures are considered in the RDC. These diagnostic criteria are proposed for clinical and epidemiological research purposes, but not necessarily for clinical practice aims (Dworkin and LeResche, 1992). In addition, RDC is the only system that considers the psychological dimensions of TMD. Its examination procedures include two axes: assessment of the physical condition (axis I) and of pain-related disability and psychological status of the patient (axis II).

The RDC product was first published in 1992 by a team headed by Samuel Dworkin and Linda LeResche (Dworkin and LeResche, 1992). It represented an interdisciplinary effort as its team comprised of recognised researchers and experts in different domains ranging from basic biological sciences to clinical dental and bio-behavioural sciences. Since they first published their original paper (Dworkin and LeResche, 1992), the RDC team have been revising and updating their work by testing the validity and reliability of several components of the physical examination and the psychological element. To further enhance the RDC, a web site[1] has been established to publish all studies related to the RDC. In addition, this site presents the RDC's history-taking questionnaire and evaluation forms in several languages[2], with the aim of using it to collect multi-national data (Yap, 2002).

[1] http://rdc-tmdinternational.org/
[2] Fourteen languages up till March 2003, and additional 3 languages presented by December 2006.

The scope of the original RDC project was multiple:

1. To critically assess the TMD diagnostic systems in use
2. To critically appraise reliability and validity of TMD examination methods
3. To standardize TMD examination procedure and measures
4. To establish research diagnostic criteria for TMD
5. To assess pain related disability and psychological status in TMD patients.

The physical component of the RDC (Axis I)

The RDC diagnostic protocol contains three major elements:

1. The evaluation forms including a history-taking questionnaire, physical examination and psychological assessment forms
2. The examiner guidelines that aim to maximize the inter-examiner reliability of the RDC examination procedure and to standardize the pain – psychological scoring process
3. The diagnostic criteria (case definitions) and algorithms allowing the diagnosis to be made from components of the RDC history and examination information.

RDC categorises TMD physical diagnoses into three groups according to the common clinical findings (Table 2.2). Other TMJ disorders, which are uncommon, or have no reliable criteria or examination methods, are excluded.

Table 2.2. Research diagnostic criteria for TMD (RDC/TMD)

GROUP I:	**Muscle disorders**
	Myofascial pain
	Myofascial pain with limited opening
GROUP II:	**Disc displacements (DD)**
	DD with reduction
	DD without reduction with limited opening
	DD without reduction without limited opening
GROUP III:	**Arthrides**
	Arthralgia
	Osteoarthritis
	Osteoarthrosis

A subject can be allocated one muscle disorder at the most, but each joint can be assigned a diagnosis from either the second or third group or both. Therefore, a number of diagnoses can be given to one person at one time, ranging from zero (no TMD diagnosis) up to five diagnoses. However, it is uncommon to assign more than three diagnoses to one patient.

Several pages in the original article describe the algorithms for the physical diagnosis. Some essential features required to make the diagnosis are mentioned in the history form and others in the examination guidelines. For these reasons, I summarised the key criteria of each category of RDC/TMD (Appendix 1) to make the diagnostic process easier. The derived form was adopted to make the diagnosis for all patients recruited in this project.

The psychosocial component of the RDC (Axis II)

Axis II concerns pain-related disability and psychological status. It consists of three scales integrated in the History Questionnaire of RDC. These are the Jaw Disability Checklist, the Symptom Checklist 90 Revised (SCL-90-R) and the Graded Chronic Pain Scale. These are covered later in the following sections.

Main advantages of the RDC system

1. The dual, physical and psychological axes of the instrument
2. The detailed examination form
3. The comprehensive history-taking from, although some questions appear superfluous to making a diagnosis
4. The examination guide that standardises the examination procedure among examiners, in particular:
    a. Patient's position during examination
    b. Patient's instructions and explanations
    c. Amount of the force used to palpate muscles and joints

d. Consideration of the vertical incisal overlap when measuring mouth opening
e. Similarly, consideration of any midline deviation in the rest position when assessing excursive movement
f. The recommended repeated measurement (e.g. evaluating mouth opening and clicks) to maximise reliability
g. The "international style" of the history form which enhances the international application of the instrument and the comparison of multi-centre data.

Main disadvantages of the RDC system

1. The complicated layout of the criteria and examination forms
2. Although the RDC team criticized other instruments for including superfluous questions and examinations which can mislead the examiner, it incorporates, itself, many items which provide no information to make a diagnostic decision. Furthermore, some of these items appear to lack reliability. For example:
   a. Assisting patients in opening their mouths
   b. The differentiation between fine and coarse crepitus
   c. The determination of three degrees of muscle tenderness: mild, moderate and severe pain
3. The examination form excludes non-masticatory muscles. Moreover, the dental examination is ignored completely. There is no recognition of any correlation between teeth, occlusion and TMD.
4. Although RDC is perhaps the most translated TMD diagnostic instrument, not all translations seem to follow scientific translation methods such as the *Forward-Backward Translation* method used for example to produce the German version (John *et al.*, 2006a). Some versions are rather based on personal efforts. In particular, the Arabic version, with which the author of this thesis is familiar, is poorly made and was clearly written by a non-native speaker. Examples of statements from this version are presented in Appendix 1. Further work is needed to produce an Arabic version using validated methodology.

2.2.4.3. The Temporomandibular Index TMI

This measure of the severity of the TMD signs was published in 2002 (Pehling *et al.*, 2002). The authors based their index on the RDC's classification, but worked to improve its examination form, to allow it to be used an outcome measure, a use for which it was not originally designed (Pehling *et al.*, 2002). The examination form of RDC was replaced by the previously published Craniomandibular Index CMI (Fricton and Schiffman, 1986) which allows for such measurement.

The new examination form (Pehling *et al.*, 2002) improves that used in the RDC, and covers almost all the information required to make an RDC diagnosis. However, the index suffers from some shortcomings such as:

1. It gives equal weights to different problems, for example, a patient with four tender muscles can obtain the same score (four points) as a patient with two tender muscles and a reproducible closing click in each joint. According to the RDC, the first subject has Myofascial pain, while no RDC diagnosis can be made for the second subject
2. The TMI also endeavours to achieve an overall score for each subject. Again, this score is clinically meaningless as it could result from any combination of conditions and makes untested assumptions regarding the impact of each condition. The improvement or deterioration of this total score is consequently of dubious value in monitoring treatment outcomes
3. The actual wording on the form is ill-defined for some items. For example, the examiner is required to differentiate between fine and coarse crepitus. This could mislead the examiner and lead to a reduction in inter-examiner reliability.

In order to overcome these shortcomings and to simplify the "crowded" form, I produced a new form based on that by Pehling *et al.* (2002). This is presented in Appendix 1.

2.2.4.4. Measuring functionality of the masticatory system

Speech, laughing, yawning, chewing and other mouth functions are known to be impaired by temporomandibular disorders (Kropmans *et al.*, 1999). Mouth function is usually assessed by means of measuring mouth opening principally in the vertical dimension but occasionally in the horizontal level, i.e. extent of excursion movements (Pehling *et al.*, 2002).

In addition to the physical measurement of mouth opening, a few instruments were developed to explore the TMD associated limitations of mouth functions. For example, Stegenga *et al.* designed a Mandibular Function Impairment Questionnaire (MFIQ) which contains 17 items (Stegenga *et al.*, 1993). Also, the RDC product has as part of its Axis II a Jaw Disability Checklist (Table 2.3) (Dworkin and LeResche, 1992).

Table 2.3. Jaw disability checklist

What activities does your present jaw problem prevent or limit you from doing?	No	Yes
Chewing	0	1
Drinking	0	1
Exercising	0	1
Eating hard food	0	1
Smiling / laughing	0	1
Sexual activity	0	1
Cleaning teeth or face	0	1
Yawning	0	1
Swallowing	0	1
Talking	0	1
Having your usual facial appearance	0	1

Measuring functioning of the masticatory system is not limited to questionnaire-based instruments. Various devices have been proposed and used to examine the masticatory system "objectively". Among these are electromyography and jaw tracking devices. The validity and reliability of these devices have not been supported, however (Mohl *et al.*, 1990; Lund *et al.*, 1995). It is beyond the scope of this review to consider these devices in more detail.

2.2.5. TMD assessment – the psychological dimension

The second dimension in the ternary biopsychosocial model of TMD (Dworkin and Massoth, 1994; Suvinen *et al.*, 2005) is the psychological dimension. This section reviews the literature for the role psychological factors play in TMD and reports the instruments that have been used to evaluate the psychological component of TMD.

There is now considerable evidence to support the importance of psychological factors in the aetiology and management of TMD. It is not yet clear, however, whether these factors cause TMD or reflect the impact of TMD on the patient (LeResche, 1997a; Suvinen *et al.*, 2005). It is recognised that the majority of TMD patients cope adequately with their symptoms, but there is a group of sufferers with psychological dysfunction who cannot cope, showing higher rates of depression, somatization and health care utilization (Dworkin and Massoth, 1994; Yap *et al.*, 2002). For example, Yap *et al.* reported that 39% of TMD patients are clinically depressed, and 55% have moderate to severe somatization (Yap *et al.*, 2002). The authors also found that severe somatization may be associated with an increase in jaw disability (Yap *et al.*, 2004).

Based on strong evidence of the role of psychological traits in TMD, some workers have argued that TMD patients should be routinely screened for psychological and psychosocial dysfunction in order to tailor treatment to their needs and to avoid failing management, often resulting in a longer chronicity of the condition (Rudy *et al.*, 1995; Turk, 1997; Dworkin *et al.*, 2002b; Suvinen *et al.*, 2005).

The Symptom Checklist 90 Revised (SCL-90-R) (Derogatis and Cleary, 1977; Hardt *et al.*, 2000) is a scale of psychological dysfunction, which measures both somatization and depression. Somatization refers to preoccupation with physical disturbance (Dworkin and LeResche, 1992). While the RDC team do not actively encourage the use of this vague term,

they favoured SCL-90-R over other measures of psychological status, such as the Beck Depression Inventory (Beck *et al.*, 1961), because it covers both somatization and depression.

Some workers have suggested that patients can be subdivided into groups based on the results of psychological tests. Butterworth and Deardorff (1987) were the first to present such a taxonomy of TMD patients. Using the SCL-90-R (Derogatis and Cleary, 1977) Butterworth and Deardorff suggested three subtypes of TMD: psychologically normal, moderately distressed, severely distressed (Butterworth and Deardorff, 1987). In 1989, Rudy *et al.* used the the Multidimensional Pain Inventory (MPI) (Kerns *et al.*, 1985) (Appendix 8) to categorise TMD patients into three subtypes: adaptive copers, interpersonally distressed and dysfunctional (Rudy *et al.*, 1989). At a later stage, the Temporomandibular Pain Dysfunction Questionnaire (Suvinen *et al.*, 1997c; Suvinen *et al.*, 1997b) was proposed and used to classify TMD cases into three subtypes: simple, intermediate and complex TMD (Suvinen *et al.*, 1997a).

2.2.6. TMD assessment – the psychosocial dimension

The psychosocial dimension of TMD is the third component of the biopsychosocial model of TMD introduced earlier (Dworkin and Massoth, 1994; Suvinen *et al.*, 2005). It is difficult to draw a clear boundary between the psychological and psychosocial dimensions. This section reports instruments that have been used to examine the impact of TMD on patients' quality of life and well-being including physical and emotional functioning. Sections 2.3.3 and 2.3.4 explore in more detail the theoretical frameworks and dimensions of quality of life and health-related quality of life with particular emphasis on chronic pain disorders.

The impact of TMD on quality of life (QOL) was first demonstrated in 1989 (Reisine and Weber, 1989). It has since received increasing recognition (Murray *et al.*, 1996; John *et al.*, 2002; Larsson *et al.*, 2004; Segù *et al.*, 2005). TMD sufferers have been found to have low

quality of life (QOL) in comparison with the normal population especially in relation to worries, sleeping, feeling tense and depressed.

Attempts to quantify the psychosocial impact of TMD have mostly relied on generic instruments as opposed to oral-specific or even TMD-specific. Workers who have used generic instruments argue that these instruments are better than specific measures in that they allow comparisons between different conditions (Di Fabio, 1998), while those in favour of condition-specific measures claim that these measures are more accurate and sensitive (Hill *et al.*, 1996; Turk *et al.*, 2003). Further exploration of this issue is presented in sections 2.3.4.2 and 2.4.3.3.

2.2.6.1. Generic instruments

Numerous standardised instruments have been used to measure pain-related disability in TMD (LeResche, 1997a). Some of these are specific to one dimension of pain, for example, perceived pain intensity or pain-related depression, while many other tools are multidimensional, which measure disability in relation to both physical and emotional function (LeResche, 1997a). Listed below are eight instruments used to measure pain-related disability in TMD. It is worth mentioning that some of the studies using these instruments did not differentiate between patients with TMD and those suffering from other orofacial pain conditions.

1. McGill Pain Questionnaire (MPQ) (Melzack, 1975) is a widely used measure of pain. It has been validated and used extensively in pain research. In addition to measuring pain MPQ measures pain-related disability (LeResche, 1997a). Both MPQ and its short form (Melzack, 1987) have been used to measure pain and related disability in a number of orofacial pain conditions including TMD (Rudy *et al.*, 2001; Van Wijk and Hoogstraten, 2002).
2. The Symptom Checklist 90 Revised (SCL-90-R) (Derogatis and Cleary, 1977; Hardt *et al.*, 2000) is a measure for both somatization and depression. It has been used as part of Axis II of the RDC (Dworkin and LeResche, 1992; Dworkin *et al.*, 2002a) but has also been used as an independent measure (Yap *et al.*, 2004).

3. The Multidimensional Pain Inventory (MPI) was first published in 1985 (Kerns *et al.*, 1985) (Appendix 8). The instrument has been validated and used extensively in pain studies including research on TMD. It is also considered "the gold standard" multidimensional self-report instrument for pain (Newton-John, 2002). Furthermore, it is recommended as one of the best two indices in pain trials (Turk *et al.*, 2006). As reported earlier, MPI has been used as a diagnostic tool to distinguish between TMD patients with different levels of psychosocial suffering (Rudy *et al.*, 1989; Dahlstrom *et al.*, 1997; Dahlstrom *et al.*, 2000) and to generate pain profiles in chronic disorders (Turk, 1997). The MPI questionnaire comprises 52 six-point items divided into three groups: *Part 1* (20 items) is related to the impact of pain on the patient's life, *Part 2* (14 items) on the support the patient receives from spouse and close relatives, and *Part 3* (18 items) on general functioning such as doing house-hold chores.
4. The Medical Outcomes Study (MOS) (Tarlov *et al.*, 1989) and its 36-item short from (SF-36) (Ware and Sherbourne, 1992) have been used in studies on TMD and orofacial pain (Kohlmann, 2002; Ta and Dionne, 2004). Di Fabio has also derived a new 17-item short form (MOS-17) to compare disabilities and health status associated with TMD to other musculoskeletal disorders (Di Fabio, 1998).
5. The Graded Chronic Pain Scale (Von Korff *et al.*, 1990) (Appendix 1) is a seven-item scale that measures perceived intensity of pain and interference of pain with general activities. It has been widely used in TMD research, especially as part of the RDC's Axis II. More recently, it has been used to produce pain profiles of TMD patients in primary health care (Forssell *et al.*, 2005) that are similar to those produced by the MPI (Rudy *et al.*, 1989).
6. The Pain Disability Index (PDI) (Pollard, 1984) has been used in orofacial pain research (Bush and Harkins, 1995; Auerbach *et al.*, 2001).

Another two generic pain-related disability indices have also been used, but in a modified form made suitable for TMD:

7. A modified version of the scale for measuring the Activities of Daily Living (ADL). This has been used in patients with craniomandibular disorders (List and Helkimo, 1995).
8. A pain-related Limitations of Daily Function in the TMD Questionnaire (LDF-TMDQ) (Sugisaki *et al.*, 2005). This questionnaire is still under development, and contains a number of instruments to cover the multidimensional nature of TMD.

The RDC team reported in their comprehensive review (Dworkin and LeResche, 1992) three of these measures: the MPI (Kerns *et al.*, 1985) (Appendix 8), the Graded Chronic Pain Scale (Von Korff *et al.*, 1990) (Appendix 1), and the pain index part of the Medical Outcomes Study MOS-PI (Tarlov *et al.*, 1989). The authors of that review did not find any of these three methods to outperform the others, but opted to use the Graded Chronic Pain Scale as part of their History Questionnaire.

2.2.6.2. Oral health-specific indices

A large number of oral health-related quality of life (OHrQOL) has been developed. These are reviewed in detail in section 2.6. Among these, the Oral Health Impact Profile (OHIP) (Slade and Spencer, 1994) is apparently the only one that has been used in the TMD population.

OHIP has been shown to be able to capture the impact of TMD and orofacial pain on sufferers' lives (Murray *et al.*, 1996; John *et al.*, 2002; Larsson *et al.*, 2004; Segù *et al.*, 2005). More recently, John *et al.* have shown a strong correlation between health-related quality of life assessed by OHIP and RDC/Axis II measures (John *et al.*, 2007).

OHIP, however, consists of 49 items, many of which are not relevant to TMD, for example, questions related to tooth sensitivity and food catching. Distilling OHIP to a TMD-specific subset of items has the potential to improve the instrument utility as an outcome measure for TMD. Section 2.6.7 explores in greater detail findings of studies that have used OHIP in TMD, and the need to obtain a short form of OHIP that is specific for TMD.

2.2.6.3. TMD-specific indices

There have been only a few attempts to develop TMD-specific instruments that address its psychosocial dimension. These are listed below:

1. The TMJ Scale (Lundeen *et al.*, 1986; Levitt and McKinney, 1994). Like the RDC, this scale was designed as a multidimensional instrument. It contains 97 items divided into three domains: physical, psychosocial and global. Although psychometric properties have been established, it has been rarely used in the TMD literature.
2. The Research Diagnostic Criteria, RDC (Dworkin and LeResche, 1992). While Axis I (physical component) of this tool was designed specifically for TMD, Axis II employs generic instruments such as the Symptom Checklist 90 Revised (SCL-90-R) (Derogatis and Cleary, 1977) and the Graded Chronic Pain Scale (Von Korff *et al.*, 1990). The social impact of TMD on quality of life, however, is not included in this assessment.
3. The Manchester Orofacial Pain Disability Scale (Aggarwal *et al.*, 2005). This instrument is principally designed to measure pain-related disability, and does not have a physical component. This scale is not designed for TMD specifically, but is targeted to all orofacial conditions. It was established from a pool of patients with orofacial pain including those consulting dental casualty and oral medicine clinics. While promising, it has not been extensively tested.

2.3. Assessing health in chronic disorders – domains of assessment

Assessing health care outcomes takes place within the wider concept of measuring health status. The traditional medical model of health determines treatment as an elimination of a disease. It considers the successful intervention the one that leads to the absence of any abnormality in the clinical examination and in the results of any complimentary physical tests (e.g. laboratory and radiography investigations). This is a limited disease-based view of ill-health and is no longer popular. Conversely, a wider paradigm of health has increasingly replaced the traditional notion, taking into account the importance of the psychosocial impact of illness on the patient's life (Allen and Locker, 1997).

The psychosocial impact gains even a greater importance when dealing with chronic or life-threatening illnesses. Here, the disease is something that the patient has to live with, and improving the quality of life becomes more important than merely prolonging survival regardless of its quality. Furthermore, it should be noted that patients with such long-term diseases will not suffer only from the disease but also from the adverse effects of treatment. Sometimes treatment makes an impact on quality of life that is worse than the disease itself. For example, chemotherapy in the management of a young lady with an invisible cancer causes her hair loss with enormous effect on her appearance and social life. This is why health bodies and regulatory agencies require evidence of intervention efficacy in the form of quality of life outcome (Johnson and Temple, 1985; National Institute for Health and Clinical Excellence, 2006).

The term "Subjective Health Status" is used in the literature to describe patient-based perception of health. It is important here, however, to distinguish this term from the clinician's subjective assessment of patients, something considered as personal, not evidence-based and therefore unreliable. Based on this definition come the terms Subjective Health Assessment and Objective Health Assessment to describe respectively patient-based and physical-based assessments of health, medical care and outcomes.

2.3.1. Rationale for the use of subjective health measurement

In the past, researchers have always called for objective measurements such as the pulse rate and range of motion. These are often characterized by good levels of reliability and sensitivity to changes. Nevertheless, many of these measures still lack content validity. In other words, the degree to which a test measures what it is meant to measure is questionable. For instance, is it the amount of mouth opening that reflects the presence of a temporomandibular disorder, or is it the patient's inability to chew the food that matters?

The deviation away from using purely objective measures is historically supported by the following arguments:

1. Measuring pain and feelings

Pain is perhaps one of the most reported symptoms of all diseases or disorders. Yet, with all the advancement of medical technology (see section 2.4), pain and other feelings cannot be reliably investigated with machines or physiological measurements. Pain is a uniquely personal experience. Furthermore, while recent advances in technology make it possible to "image pain" (Porro, 2003), this is still research work and technically impractical in daily clinical practice. Therefore, until further development of these techniques becomes reasonably obtainable, asking patients about their feelings is more practicable than measuring it with sophisticated, expensive and time-demanding devices. Moreover, patients' self report of pain is considered the gold standard measure in pain research (LeResche, 1997a).

2. Problems with the validity of objective measures

These measures reflect the biological status of the body but might be inconsistent with the person's perception of this condition. There are various circumstances where patients may demonstrate positive objective signs of illness, while they are unaware of their health problems or consider the symptoms they experience as part of everyday life. For instance, the

prevalence of a disease, such as cholera, in a specific area would makes its inhabitants familiar with the disease and consider it as normal. On the other hand, clinical evidence and experience describe patients who complain of sickness, but no objective proof supports their complaints (Jenkinson, 1994). Furthermore, studies which have investigated the association between clinical and subjective (patient-based) measurement of health have found that this relation is weak (Locker and Slade, 1994).

3. The invalidity of the "health care use" and other societal indicators

"Health care use" indicators have been used to assess treatment outcomes and the need for further treatment. They are not a good alternative to objective measures, however, as they do not solely reflect the patient's need for further treatment. They are rather influenced by the care provider's policies and resources (Bowling, 2005). Other societal tools do assess the impact of illness on individuals, including days of restricted activity and work or school absenteeism. While these indicators are useful for demonstrating the burden of a given disease on the population, they are of doubtful benefit when assessing the impact of this disease at an individual-level (Slade, 2002). On the one hand, they are not suitable for those who are not in the workforce. On the other hand, people with even severe levels of discomfort may still attend work. This is again influenced by the policies in the organisations where they work.

4. Weakness of the professionals' assessment

Studies show that health professionals are poor judges of the impact of illness and treatments on the patient's everyday life (Sprangers and Aaronson, 1992). This gap between health care providers and seekers can be reduced by incorporating the latter's views of their conditions and how these change with treatment (Gill and Feinstein, 1994).

5. The cost-benefit analysis of health care

The call for evidence-based medicine has arisen because of the controversy over different treatment plans for the same disease, both internationally and nationally (Jenkinson, 1994). Moreover, the discrepancy between the rising costs of health care and the escalating levels of mortality, morbidity and disability make health care funders require more evidence on the benefit of the medical costs they are paying (Corson *et al.*, 1999). Consequently, they no longer accept the medical experts' views of health care and its outcome; instead they want the "consumer's perspective" in the issue. For this purpose, a vast number of patient-based questionnaires have been developed. These assess and compare the gained quality of life using different interventions. Examples of these instruments is the *Quality-Adjusted Life Years (QALYs)* (Nord, 1992). Health care regulators such as the National Institute for Health and Clinical Excellence in the U.K., request clinicians to base their treatment decisions on these types of questionnaires, known as utility measures (National Institute for Health and Clinical Excellence, 2006).

6. To comply with the World Heath Organization (WHO)'s definition of health

In 1946 "WHO" stated that "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (World Health Organization, 1946). This statement highlighted the importance of incorporating the concepts of function, socio-psychological contentment and subjective perceptions of satisfaction in any generic concept of health. The term "health-related quality of life" emerged as a concept that encompasses the spirit of the WHO definition of health by incorporating both personal health status and social well-being in assessing the health of individuals and populations (Guyatt *et al.*, 1993). The notion of subjective health assessment developed simultaneously. Since that time, health-related quality of life data have seen an increase in publication. It has been documented, for example, that physicians and researchers have an increasing interest in utilizing the patient's view point in planning treatments and monitoring their outcomes (Geigle and Jones, 1990). Consequently, measures of health status that fail to incorporate one of WHO's dimensions are subject to negative evaluation (Bowling, 2005).

2.3.2. Applications of subjective health measurement

Subjective (patient-based) health measures have been used for different purposes including the following (Fitzpatrick *et al.*, 1992a; Gift and Atchison, 1995; Corson *et al.*, 1999):

For clinical applications and researches

1. Assessing clinical trials' outcomes

Interventions designed to improve patients' quality of life, rather than to increase life expectancy, are increasingly interested in evaluating treatment outcome using patients' views of their daily lives, instead of the physicians' views on their behalf. An example of this type of treatment is joint replacement, where the principal aim is to reduce the pain and restore the patient's ability to carry out everyday activities. Quality of life measures are most suitable for phases II and III of clinical trials, which aim to assess the effect of treatment on the participants (Osaba, 2005).

2. Providing physicians with a comprehensive picture about their patients' status

As clinicians endeavour to treat the patient and not the disease, patient's perception of illness and its changes during the course of treatment are indispensable auxiliary tools in medical practice. Furthermore, understanding patients' needs and incorporating them in making decisions about their illness management enhances their co-operation and response to treatment instructions (Bowling, 2005).

For resource allocation decision-making

3. Rationalizing the cost-utility balance of medical interventions

Limited resources along with rising health care costs make health care funders and authorities adopt the quality of the patient's life as a basis in making rationing decisions. Therefore, the

impact of interventions on patients' health status and quality of life has been used to justify the choice of treatment (Fitzpatrick *et al.*, 1992a).

4. Population studies

Subjective health assessment has been successfully used to inform health authorities about the population health status and needs (Kelly, 2000; McGrath and Bedi, 2003a). Studies might involve the whole population or a population of specific characteristic, e.g. TMD patients (Dworkin *et al.*, 1990; Dworkin and LeResche, 1992). Data obtained by these sorts of studies can help as normative references for comparison between definite groups, e.g. how do diabetics' scores differ from healthy individuals.

2.3.3. Health components and indicators

Health, and similarly health-related quality of life, has different components. There is, however, a lack of consensus amongst researchers over the definitions and measurements of these components, and these are often used interchangeably (Bowling, 2005). The most widely-used concepts in the literature are: functional ability, positive health, social health, subjective well-being, and quality of life. The literature records an increase in the use of broad health status and health related quality of life indicators over the past 20 years (Bowling, 2005).

There is also a large emphasis on establishing indicators that measure health positively rather than negatively. In other words, indices that assess healthiness and good quality of life as opposed to deterioration of health and the consequent impact on quality of life (Bowling, 2005). Realistically, with the lack of an operational definition of health, it is difficult to establish such "positive" measures, and the negative approach still dominates most current health indicators. There is a need for broad health status scales that adopt the positive concept and measure health as a continuum rather than measuring the deviation away from the healthy status (Bowling, 2005).

The focus of this review is on the measurement of patient-based health-related quality of life. Numerous batteries of instruments have been developed to assess health both objectively and subjectively, i.e. from the clinician's and patient's perspectives. But it is beyond the scope of this project to review clinical indicators and non-biological indicators such as length of hospital stay, time to return to work and hospital readmission rates.

The measurement of patient-based health-related quality of life (HRQoL) is typically achieved by evaluating the different domains of health including physical function, mental health, emotional state, social well-being and life satisfaction (Bowling, 2005). Patient-based HRQoL measures are built on conceptual frameworks that encompass these domains (e.g. physical functioning) and test them through their dimensions (e.g. mobility). These conceptual frameworks have been developed in oral health (Locker, 1988; John *et al.*, 2004a; Bagewitz *et al.*, 2005).

2.3.4. Core outcome domains for chronic pain research

This section reviews the literature on the measurement of health in chronic pain conditions within the broader picture of health and its domains. The focus here is on the definition of pain, the subjective nature of pain and dimensions of pain.

The International Association for the Study of Pain (IASP) defines pain as "an unpleasant sensory and emotional experience associated with actual or potential damage, or described in terms of such damage"(Merskey *et al.*, 1994). Pain has often been regarded as a passive warning signal of an underlying disease process. However, it has been increasingly argued that persistent pain, which lasts beyond the expected period of healing, is a disease entity by itself (Siddall and Cousins, 2004). These last authors argue that accumulated evidence suggests that persistent pain has its own pathology, signs and symptoms and meet all the criteria to be considered and treated as a disease entity. They believe that the sole focus on removing the primary underlying pathology that triggered the persistent pain without recognising the secondary pathology and consequences simply fail to obtain satisfactory outcome. In this respect, treatment decisions as well as outcome assessment should address the many consequences of chronic pain (Siddall and Cousins, 2004).

Undoubtedly, the personal experience is a core issue in measuring pain, and despite the technological advancement that allows an "objective assessment" of pain (section 2.4), it is widely believed that patients' self-report of the pain experience is the gold standard in pain research (LeResche, 1997a).

Dimensions of pain

IASP's definition of pain recognises the personal nature of pain but highlights also its multidimensional nature. As with the concept of health, there has been long debate about the dimensions and components of pain.

In 1968, Melzack and Casey suggested a three dimensional model of pain: sensory (discriminative), motivational (affective: motivating the organism into activity aimed at stopping the pain), and cognitive (evaluative) (Turk and Melzack, 2001). In this model, the three dimensions interact with one another to inform the brain about the noxious stimulus and to motivate the body toward acting against this effect based on past experience and knowledge about this stimulus.

Other researchers (LeResche, 1997a; Newton-John, 2002) have suggested a wider multidimensional concept of chronic pain, which incorporates the individual perception of pain as well as the psycho-social dimension of the disease (chronic pain). These authors recommend that the subjective (patient-based) assessment of pain should consist of a battery of the following domains:

1. Perception of pain: intensity, quality, duration and location
2. Physical functioning: localised to the diseased area; generic concerns about the daily activities; and social role functioning
3. Emotional functioning: depression, anxiety, anger
4. Cognitive appraisal of pain: pain-coping; and beliefs about pain e.g. ability to control pain and catastrophising
5. Behavioural pain including expressive behaviour, pain-related limited activities and health care usage
6. Personality
7. Family support
8. Improvement and satisfaction with treatment.

More recently a group of 27 authors from academia, governmental agencies and the pharmaceutical industry launched an initiative to establish a consensus about the core domains of chronic pain that must be considered when assessing outcomes in clinical trials

(Turk *et al.*, 2003). The project, which was called IMMPACT (Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials), suggests six domains of chronic pain treatment outcomes. These are:

1. Pain (i.e. perception of pain)
2. Physical functioning
3. Emotional functioning
4. The individual's ratings of improvement and satisfaction with treatment
5. Symptoms and adverse events
6. The participant's disposition (adherence to the treatment regimen).

While this work is an advanced, multi-authored project, and is close to the views of both LeResche and Newton-John (above), it just reflects that there is still a lack of consensus over pain domains, and that these are still been used interchangeably.

In relation to quality of life in chronic pain patients, the IMMPACT team found in their review that only two main domains in HRQOL (physical functioning and emotional functioning) gain sufficient consensus to be considered in the assessment of clinical trials (Turk *et al.*, 2003). This conclusion supports results of other studies (De Gagne *et al.*, 1995; Holroyd *et al.*, 1999; John *et al.*, 2004a) where factor analysis was used to identify the main domains of the pain's impact on quality of life.

In the light of these recommendations, it is necessary to consider the perception of pain, physical functioning and emotional functioning in more detail.

2.3.4.1. Perception of pain

Measuring patients' perception of pain comprises the examination of its intensity, quality, frequency, duration and location (LeResche, 1997a). The change in intensity is perhaps the most frequently used measurement of treatment outcomes. In contrast, while pain quality,

duration and location are essential to making a diagnosis, they are rarely used to measure treatment outcomes (LeResche, 1997a).

Self-report instruments are considered the gold standard in pain studies (LeResche, 1997a). The most common of these are Visual Analogue Scales (VAS), the McGill Pain Questionnaire (MPQ) and its short form The Descriptor Differential Scale (Turk and Melzack, 2001). These instruments are discussed in more detail in section 2.4.4.4.

2.3.4.2. Physical functioning

One of the oldest and most common methods of assessing treatment outcomes is the patient's ability to perform daily tasks. Functional ability is considered "the domain which has the most face validity for the patients themselves as it points to the ways in which the pain has changed their daily lives" (Newton-John, 2002).

Several models of disability exist; among these is the widely-used WHO's classification of functional disability: the International Classification of Impairments, Disabilities and Handicaps (ICIDH 1980) (World Health Organization, 1980). This model defined three concepts or degrees of disability: impairment, disability and handicap. An example of this model is shown in Table 2.4.

Table 2.4. An example of WHO's classification of functional disability*

Disease or Disorder	Impairment	Disability	Handicap
Blindness	Vision	Seeing	Orientation

*** From (Bowling, 2005)**

In this model, impairment relates to the biological function of the body, whereas disability focuses on the effect on activities and performance resulting from the impairment. Handicap relates to the social consequences of impairment and disability.

In 2001, WHO developed a broader classification of functioning and disability: the International Classification of Functioning, Disability and Health (ICF) (World Health Organization, 2006) which adopts a positive approach to health. In other words, the new model does not give the disability a formal definition. It considers disability as a limitation of function and focuses on the different domains of function (body function, activities and participation in life situations) (Bowling, 2005).

Scales measuring this broader view of functioning and ability are available in two types: generic and condition-specific (section 2.4.3.42.4.3.3). Generic measurements such as the widely-used short form 36 (or SF-36) (Ware and Sherbourne, 1992; RAND, 2006) assess whole body function and have been used in a large number of conditions including TMD (see section 2.2.6.1). Condition-specific instruments, on the contrary, are designed to evaluate the impact of a specific condition such as TMD, on quality of life. Examples of these scales are presented in sections 2.2.6.3 and 2.4.3.3.

It is argued, on the one hand, that generic instruments may not capture changes in the disorder itself (Hill *et al.*, 1996; Turk *et al.*, 2003), and therefore disease-specific measures are more likely to reveal these changes in much more detail. On the other hand, these specific measures may fail to reflect on the overall impact of the disease on QOL, where generic instruments will be useful and needed. The use of generic and specific measures in combination was then recommended to achieve the best assessment in clinical trials (Patrick and Deyo, 1989; Turk *et al.*, 2003). It has been also suggested that engaging in social activities is of less importance in clinical trials and thus social role functioning is not a core domain in treatment outcome measures (Turk *et al.*, 2003).

2.3.4.3. Emotional functioning

While cognitive aspects of pain have rarely been used as primary outcomes, many studies aimed at modifying pain-related cognitions for chronic pain problems (LeResche, 1997a). Moreover, different instruments have been developed to assess pain-coping including the Coping Strategies Questionnaire (Rosenstiel and Keefe, 1983).

The assessment of behavioural pain is a multi-axial model itself. It comprises assessing expressive behaviour such as facial expression of pain; pain-related modifications of activities e.g. guarded movement; and health care usage both for medication and clinical attendance (LeResche, 1997a).

2.4. Assessing health in chronic disorders – methods of assessment

As noted earlier, assessment of health is performed using objective or subjective measures. After a brief summary of objective measures, this section reviews the literature on subjective (patient-based) methods of health assessment. It covers their types, uses, advantages and disadvantages, and properties. Examples of some pain measures are presented at the end of this section.

Objective measures of pain include physical and biological tests, such as sweating, heart rate, blood pressure, respiration, oxygen saturation and endorphins levels. Recent developments in imaging techniques have also made it possible to "image pain" (Schweinhardt *et al.*, 2006). In other words, to measure changes in the Central Nervous System (CNS) during pain. Most of these measures are more appropriate for acute pain and experimental pain studies than chronic pain research. Moreover, the validity and reliability of many of them are still under investigation. Other objective measures include behavioural and function measures, but it is beyond the scope of this review to explore the vast variety of these methods in more detail.

Patient-based reports of health can be obtained either through interviews or by means of self-report (or self-administered) instruments including diaries, questionnaires and scales. While the interviews yield qualitative data, self-report measures are capable of producing both quantitative and qualitative data.

2.4.1. Interviews

Interviewing patients provides a rich source of data for researchers interested in the patient's perception and assessment of health. These, along with other methods such as *Focus groups*, *Delphi Panels*, and *Observation* are known as *Qualitative Methods*, or *Qualitative Research*. In such cases, the data are not numerical, but textual, usually in the form of direct quotes, ideas or descriptions. Qualitative research is very important in pain research because of the personal nature of pain and its complexity, which can sometimes be difficult to be explored using quantitative methods. In particular qualitative research is a very useful tool to understand the lived experience of pain.

2.4.1.1. Qualitative research

Qualitative methods play an important role in clinical research, in particular in social medical research. Although qualitative research is not as common as quantitative research, both types are linked with one another, and often findings from qualitative research establish the foundation for quantitative research. For example, patient-based health questionnaires are often derived from statements given by patients, which are then transformed into questions. An example of this approach is the derivation of the Oral Health Impact Profile (OHIP) (Locker and Slade, 1993; John *et al.*, 2002).

Qualitative research is essentially exploratory. It aims to understand and explain a particular phenomenon using questions like "what?", "why?" and "how?" as opposed to "how many?" or "how frequently?" often used in quantitative research (Gantley *et al.*, 1999).

In health research, qualitative methods are often used for three purposes (Gantley *et al.*, 1999):

1. To explain the process (of an illness or particular course of action for example)

2. To understand ideas and beliefs of the participants (for example behavioural attitudes in a given community)

3. To generate a hypothesis or a theory about a specific phenomenon.

2.4.1.2. Advantages and disadvantages of qualitative research

The differences between the quantitative and qualitative approaches are important, and an appreciation of the strengths and weaknesses of qualitative research is essential. A summary of these is provided in Table 2.5.

Table 2.5. Summary of strengths and weaknesses of qualitative research*

Strengths	• It is possible explore further the meanings that the participant attaches to a specific answer • This research is open ended: participants' own priorities are allowed to lead the data collection • Small samples allow rich and detailed data to be obtained

Weaknesses	• Data is context specific and cannot be generalised to other contexts
	• The researcher has an impact on the data:
	  ○ By influencing the respondent's answers
	  ○ Through personal and inaccurate interpretation of the data
	• Sensitive topics are difficult to discuss; participants may not provide the most accurate information
	• Small samples are used

*** From (Britten, 1995; Gantley *et al.*, 1999)**

Qualitative methods such as interviews have other disadvantages, which have to be taken into consideration. Among these is the cost for recruiting skilled interviewers, the expense of commissioning data transcription, and if necessary, the costs and skills required to use special data analysis software packages. In this section I am going to limit the review only to interviews, as these are the most relevant techniques to this thesis.

2.4.1.3. Types of interviews

Interviews are categorised under three groups (Britten, 1995):

1. **Structured interviews.** In this form of interview, also known as *Quantitative Interviews* (Britten, 1995) trained interviewers administer pre-designed questionnaires which have

structured or fixed-choice questions such as "How is your health: excellent, good, fair, or poor?". The responses have numerical codes and are subject to statistical analysis, thus this form is considered a quantitative approach.

2. **Semi-structured interviews.** Here a flexible framework (a topic guide) is provided to the interviewer, which defines the area of interest but allows for further exploration with the interviewee. Often open-ended questions are used in this type that permits the interviewee to give more details. For example: "What do you think good health is?"
3. **In-depth interviews.** These interviews are less structured than the previous. They often cover a few topics but in much greater detail. They start with defining the topic of the interview (research) but the interviewer poses questions based on what the interviewee has said. For example, the opening question might be: "Can you tell me about your own health experiences and what you think of your health?"

The term *Qualitative Interviews* is often used interchangeably with "interviews" in the literature. However, it refers to semi-structured and in-depth interviews. This term is used in this thesis to describe these two types of interviews.

2.4.1.4. Sampling in qualitative interviews

Unlike quantitative studies, statistical representativeness and numerical generalisation are not normally sought in qualitative research. Instead, conceptual analysis is more important (Fitzpatrick and Boulton, 1994; Mays and Pope, 1995). Sampling strategies are determined by the purpose of the interview (Britten, 1995), which "allow the researcher to claim either that a full range of opinions or experiences have been investigated or that the data have provided the grounds for theoretical development" (Gantley *et al.*, 1999). These sampling strategies are known as *Theoretical Sampling* (Fitzpatrick and Boulton, 1994), *Purposive Sampling* or *Purposeful Sampling* (Gantley *et al.*, 1999). Compared to the *Random Probability Sampling*

often used in quantitative research, the sample size is often small in qualitative research. As noted in Table 2.5, the largest qualitative studies may recruit only 50 subjects (Britten, 1995).

It should be noted, however, that there is no reason why random probability sampling cannot be used in qualitative research (Mays and Pope, 1995). For example, a random sample of patients could be interviewed to explore their satisfaction about a given treatment.

2.4.1.5. Considerations in data collection in qualitative interviews

There are certain basic but important issues that have to be taken into consideration in interview-based research. These include:

1. The interview settings have to be convenient for both the conduct of the interview and for the interviewee. For example, the interview should be conducted in an environment that does not allow for external interruption or noise, or for overhearing what the interviewee is saying. They also have to take place at a time that suits the interviewee (Britten, 1995).
2. The immense volume of data yielded during interviews makes it difficult for the investigator to take accurate notes of what the interviewee has said. This might also interfere with the interviewing process. For these reasons audio recording of the interviews is preferred (Britten, 1995). This, however, brings certain ethical issues, including the possibility of recognising the voice of the interviewee, and how and for how long audio records should be kept so confidentiality is maintained.

Issues such as confidentiality and anonymising of the study participants and bias in the study are common amongst all methods of patient-based assessment of health. These are further explored in section 2.4.6.

2.4.1.6. Analysing qualitative interviews data

In contrast to quantitative analysis that involves manipulation of numerical data, qualitative analysis consists of manipulation of concepts. Conceptual analysis involves identifying issues

of interest to the study participants and developing concepts of underlying patterns (Fitzpatrick and Boulton, 1994).

Different approaches are available to analyse data from qualitative interviews including: the *Analytic induction*, the *Framework approach*, *Grounded theory* and the *Narrative analysis* (Gantley *et al.*, 1999). Each method has its advantages and disadvantages, and specific indications. The common denominator among these methods is the iterative nature of the analysis (Bryman and Burgess, 1994). In other words, the investigator starts analysing the data (first interviews) by identifying the main issues emerging in the interviews. He or she tries then to come up with a hypothetical explanation (concept) for the investigated phenomenon. As more data are collected, the investigator may need to refine the developing concepts or add new subjects to the study sample to further investigate some issues. This involves going back and forth between developing concepts and the raw data emerging from each new interview (Bryman and Burgess, 1994; Gantley *et al.*, 1999).

Computer-assisted qualitative data analysis software (CAQDAS)

Technological advancements have made it possible for qualitative researchers to benefit from computers in analysing qualitative data. A large number of data analysis software packages are available for different purposes and stages in the qualitative data analysis (Murphy *et al.*, 1998). These include: text retrievers, textbase managers, code and retrieve programmes, code-based theory builders, and conceptual network builders. Examples of these packages are: Ethnograph, ATLAS/ti, and NUD*IST.

The major role these packages play in qualitative research is facilitating handling qualitative data sets and improving the rigour of the analysis by helping the search for falsifying evidence (Murphy *et al.*, 1998). Despite these major advantages, it should be noted that qualitative data analysis is not an algorithmic process. Therefore, these computerised

analysis packages can only facilitate the organisation of data. They do not carry out qualitative analysis on behalf of the human analyst (Murphy *et al.*, 1998).

2.4.1.7. Rigour in qualitative research

Qualitative research is sometimes criticised for lacking of rigour methodology. Mays and Pope (1995) have listed three sources for this criticism:

1. Some accused qualitative research of being merely a collection of anecdote and personal impressions that is strongly subject to the researcher's bias
2. Others argue that this type of research lacks reproducibility. In other words, there is no guarantee that different researchers would be able to reach the same conclusions
3. Qualitative research is also criticised for lacking generalisability.

Mays and Pope (1995) argue that qualitative and quantitative researches are not of different types, and that basic strategies governing the quality of quantitative research should be employed in qualitative research to ensure rigour. These strategies include: systematic and self conscious design, data collection, interpretation and communication (Mays and Pope, 1995). In the authors' opinion, it is not the qualitative research that should be criticised; but rather qualitative workers who fail to communicate their assumptions and methods.

2.4.2. Diaries

A diary (also called *Health Diary*) is a method of self-report assessment that is based on repeated momentary measurements. It asks patients to repeatedly record their experiences of the condition at the moment of assessment or over a brief time period (Stone and Shiffman, 2002). Unlike other self-report instruments that rely on retrospective assessment of the condition, the diary provides what is called an Ecological Momentary Assessment (EMA) (Stone and Shiffman, 1994). In other words, it captures the particulars of the experience as it is lived by the patient rather than relying on recall of past experience (Stone and Shiffman, 2002).

Since the 1940s, diary instruments have been used extensively in social, psychological and physiological research (Verbrugge, 1980; Bolger *et al.*, 2003). Examples from medicine include research on chronic pain, mood, asthma, coronary heart diseases, diabetes and cost-effectiveness (Goossens *et al.*, 2000; Gendreau *et al.*, 2003; Aaron *et al.*, 2004). Diaries were also employed in TMD research (Forssell *et al.*, 1985; LeResche *et al.*, 2003; Aaron *et al.*, 2004; Wassell *et al.*, 2004; Wig *et al.*, 2004).

Diary methods are suitable for research that examines in detail the personal experience of the condition. These research types are summarised in three categories (Bolger *et al.*, 2003; Gendreau *et al.*, 2003):

1. Obtaining an aggregate value of the experience over time
2. Exploring temporal patterns of changes in the experience, i.e. changes with time
3. Examining factors affecting changes in the experience.

For each type of topic, diary data can be examined on the person-level and for between-person variability. For example, chronic pain diaries help:

1. To obtain an aggregate value for the pain experienced by each patient (person-level) and to compare between patients (between-person variability)
2. To explore pattern of changes in pain level with time (within-person variability) and to compare different patterns of changes (between-person variability)
3. To conduct a causal analysis of within-person changes over time, and individual differences in these changes.

2.4.2.1. Advantages and disadvantages

It is important to understand and consider both the advantages and disadvantages of diary methods when designing the research protocol.

The most important advantages of diaries are the quantity and quality of generated data. These are extensive and more accurate than those obtained with traditional survey or interview methods. Diaries "reduce the likelihood of retrospection encountered in questionnaires and interviews by minimizing time elapsed between an experience and the account of this experience" (Bolger *et al.*, 2003; Gendreau *et al.*, 2003). Studies of pain show that aggregate values for discrete diary data are better than retrospective summary measures that rely on recall (Stone *et al.*, 2000). Retrospection leads to a bias that results from emphasising recent incidents. This type of bias is called "*State-Congruent Recall*" (Bower, 1981). Research on pain shows that pain experiences are heavily affected by numerous factors, such as respondents' current emotional state, level of anxiety and underlying disease processes (Eich *et al.*, 1985). Patients often tend to give more weight to a combination of peak-pain and pain proximal to the recall ratings than reflecting on a simple average of all momentary pain reports (Redelmeier and Kahneman, 1996; Stone *et al.*, 2000).

On the other hand, diaries place a remarkable burden both on the participant and on the researcher. Participants suffer the inconvenience of completing the diary repeatedly and sometimes at inconvenient times (as is the case with random-scheduled diaries – see below).

The researcher needs to train the participants and establish their understanding and co-operation. In addition, the high volume of data requires more work to insert, verify and analyse these data. Furthermore, to guarantee participants' compliance, these need to be prompted and "monitored" constantly. For example, a diary study on TMD patients (LeResche *et al.*, 2003) prompted the participants with regular telephone calls and with a total of $125 (given in instalments at different stages in the study).

Other problems include the participant forgetting, being unable, or even unwilling to complete the diary on time, leaving it with missing entries or fabricated retrospective records (Stone *et al.*, 2002). The participant may also wish not to complete the diaries (at least on time) when personal or sensitive matters are the subject of the diaries, and when these can be accessed by others (Bolger *et al.*, 2003).

Finally, some argue that completing a self-assessment diary itself has an effect on the participants and on treatment outcomes, but this issue is controversial (Cruise *et al.*, 1996; Aaron *et al.*, 2005).

2.4.2.2. Types of diary designs

Diary protocols can be classified into two main types: *Time-Based* (*Time-Contingent*) and *Event-Based* (*Event-Contingent*) (Bolger *et al.*, 2003).

Time-based protocols require patients completing the diary at predetermined fixed intervals (*Interval-Contingent*) or when prompted by a signalling device (*Signal-Contingent*). The prompting device can be programmed to generate signals at fixed intervals, random intervals or a combination of both. The fixed intervals are usually of hourly or daily frequency, e.g. patients are asked to report their pain each hour or at the end of the day. Selection of the schedule (fixed or variable) and of intervals is based on the phenomenon being investigated. But generally, a rational decision is needed as to how long the interval should be between assessments and when these assessments should take place.

In this context, signal-contingent protocols have the advantage of allowing the researcher to select times for assessment. In other words, much as with sampling design for subjects, sampling of moments can be designed to address the particular research question (Gendreau *et al.*, 2003).

Time-based protocols are most suitable for studies of within-person changes in ongoing processes. Event-based protocols, on the other hand, are most suited for the study of rare and isolated events. In these designs, patients have to report at every instant they have experienced the phenomenon of interest, e.g. only when they get pain. This design overcomes the drawback of the time-based one which can miss these isolated events with its restricted schedule. A clear definition of the triggering event is essential to ensure that the subject does not report unwanted incidents or miss those required. A combination of both time- and event-based protocols is also possible (Bolger *et al.*, 2003; Gendreau *et al.*, 2003).

2.4.2.3. Types of diary technology

The first and simplest type comprises *Paper and Pencil Diaries*. With its simplicity and convenience for lay people, this type has several shortcomings (section 2.4.2.1), which were addressed and partly overcome by new advancements in technology. *Augmented Paper Diaries*, for example, help to solve the problem of forgetfulness to complete the paper diary and the consequent problems of dealing with missing data (Broderick *et al.*, 2003). This is achieved by alerting respondents to complete the diary, usually by means of electronic devices such as pagers that signal at predetermined times. While this technology maintains the same simplicity of paper diaries and overcomes one of its disadvantages, it is costly and can be disruptive of participants' routines. Furthermore, while it eliminates the problem of participants forgetting to complete the diaries, it does not guarantee their compliance once they remember (Broderick *et al.*, 2003).

Over the last decade, *Electronic Diaries* have appeared, which yield significant improvements. These are handheld computers equipped with diary programmes. They do not merely prompt participants, but can also monitor their compliance by recording entry time, thus identifying "false" late entries (Stone *et al.*, 2002; McCracken *et al.*, 2005). For example, in a study on patients' compliance, this was found to be 94% for electronic diaries in comparison to 11% for paper diaries (Stone *et al.*, 2002). These equipments can also be programmed to accept entries within a margin of time, thus preventing entries being made retrospectively.

In addition, electronic diaries have the flexibility of posing the research questions with different presentations. They can change questions' contents and sequence according to changes in the condition, thus avoiding unnecessary repetition and cumbersomeness. Furthermore, these diaries are a means of data collection themselves. Data inserted by participants can be transferred directly to the researcher's computer, hence saving time and the error of data insertion at later stages.

Undoubtedly, these technological innovations come at a cost and have limitations. First, the costs incurred by buying and maintaining these devices, designing and installing the programmes, and training the participants. Other limitations include participants' computer literacy and worries about expensive computer equipments being stolen. Open-ended questions can be problematic with these devices as response entry becomes difficult (Bolger *et al.*, 2003; Gendreau *et al.*, 2003).

Newly emerging technologies in diaries include voice recording and recognition which allow for open-ended questions to be asked and answered by voice. In addition, Web-based questionnaires (both traditional and mobile-accessed) enable interactive contacts between the researcher and respondents with an increased compliance. Finally, future devices are expected to record both self-report, biological and physiological data (Bolger *et al.*, 2003).

2.4.3. Questionnaires

Questionnaires are important tools in the measurement of patients' perception of health, health care, and health economics. Extensive research has been undertaken to develop and validate these instruments in health research. By definition, a questionnaire is "an instrument consisting of a series of questions and/or attitude-opinion statements designed to elicit responses which can be converted into measures of the variable under investigation" (Franklin and Osborne, 1971) cited by McColl *et al.* (2001).

2.4.3.1. Modes of administration

There are four major modes of administration for questionnaires: self-completion questionnaires, face-to-face interviews, telephone interviews and computer-assisted questionnaires. Table 2.6 lists major advantages and disadvantages of these methods.

Self-completion questionnaires are the most widely used in health research as they are cheaper and require less specialised resources than interviewing. They are criticised, however, for being limited to literate respondents and for failing to provide explanations to the respondents (McColl *et al.*, 2000). Hence, they are most suitable for collecting factual data such as age. They are subject to error when used to measure attitudes and beliefs where the interpretation of the wording of the question might play a role (Bowling, 2002). On the other hand, questionnaires that are facilitated by an interviewer are also prone to bias (Bowling, 2002). The results might be affected by the interviewer's explanation. Also, lack of anonymity may influence response when it comes to sensitive topics such as personal issues or opinions about treatments. In addition, they are time and resource consuming.

Self-completion questionnaires can be administered directly in paper or electronic form, or sent to the population of interest by post. Electronic questionnaires are now increasingly used (Yap *et al.*, 2001; Couper, 2005). They have some advantages over the paper form

including the elimination of the data entry stage removing an essential source of potential error. Nevertheless, they require the participants to be computer literate and arguably may incur additional costs to the research (Yap *et al.*, 2001).

Table 2.6. Advantages and disadvantages of the main modes of questionnaire administration*

	Self-Administered				*Interviews*	
	Mailed	*On-Site Paper-and-Pencil or Disk*	*On-Disk (on-site or mailed)*	*Interactive*	*Telephone*	*In-Person*
Characteristics	Respondents use paper and pencil	Respondents use paper and pencil or respond directly into the computer	Respondent answers directly onto the computer disk	Respondent uses a computer that is linked into a network or the Internet	Can be done with written script or be computer assisted	Can be done with a written script or be computer assisted
Advantages	Can reach large geographic areas	Information is obtained immediately If supervisor is present, confusing survey questions can be clarified as needed	Can reach large geographic areas Order of questions can be preprogrammed Only "legal" answers are accepted	Can reach the world almost instantly Can give respondent an explanation of unfamiliar words and help with difficult questions Data can be automatically entered and analyzed	Can explore answers with respondents Can assist respondent with unfamiliar words	Same as telephone
Disadvantages	Need motivated respondents to return survey Respondents must be able to read, see, and write	Limited to responses from just those who are on site or can get there Respondents must be able to read, see, and write	Computer may quit Respondent must be willing to use the computer to complete a survey	Need access to system (like Internet) Respondent must be willing to use the computer to complete the survey System can go down	Need trained interviewers Need to make sure respondent is home If using computer-assisted interviews, will need technical expertise to program them	Need trained interviewers Must find a suitable place to conduct interview
Special needs	Up-to-date address list Follow-up mailings Incentives	Space and privacy for respondent to complete the survey	Programmer	Software engineer or programmer	Up-to-date phone numbers Schedule for reaching respondents May need a sampling expert for random digit dialing	If on site, need space and privacy May be difficult or dangerous to go to person's home
Costs	Printing, paper, envelopes, stamps, incentives	Printing, paper, incentives, survey supervisor, and possibly space for respondent to work	Computer, programmer, incentives	Mainly technical (e.g., software engineer), incentives	Training, incentives, telephones and telephone charges, computers and technical expertise, sampling expert	Training, space, travel, incentives

*** From (Fink and Kosecoff, 1998)**

Postal questionnaires have the advantage of being easier to administer on a large scale. In contrast to directly administered questionnaires, however, they suffer from a lower response rate and arguably higher costs. They also do not guarantee that the person who completed them is the one they are claiming to be. This is important when the help and opinions of others might affect the results. Furthermore, the researcher has no control over the completion of the questionnaire in areas where the questions' order matters.

The focus of this section is on self-administered questionnaires, and in particular the "Pencil & Paper" form. The section uses as a source of information a comprehensive systematic review of questionnaires design and use in health research by McColl *et al.* (2001). Interviewer-administered questionnaires have been previously covered in the section about interviews (section 2.4.1). Ultimately, none of questionnaires' modes of administration shows superiority in all criteria. Choice of a particular mode is mainly based on the circumstances of the research (i.e. study topic; volume and complexity of data to be collected) and on the available resources (McColl *et al.*, 2001).

2.4.3.2. Questionnaire designs

Unlike interviews where open questions are often used to allow the respondent and the interviewer to expand more on particular points, questionnaires use usually a fixed frame of questions with pre-set response choices (Bowling, 2002). The strength of questionnaires comes from the fact that they produce unambiguous answers, and yield quantitative data that are relatively easy to collect and analyse. On the other hand, pre-set choices limit the respondents and might force them to choose an answer that does not exactly represent their views.

Question wording

Posing the question is an art in itself. It has an important role in optimising completion rates and providing the most accurate information. To begin with, simple, jargon-free language

should be used. The questions should also avoid words with more than one meaning, and leading statements. Long questions and double-barrelled questions (those with "and / or") are best avoided (McColl *et al.*, 2001).

In their systematic review of the design and use of questionnaires in health studies, McColl *et al.* (2001) suggested that the question should be accurate, avoid negatively phrased attitudinal sentences, and forgo the combination of direct and indirect questions in the same questionnaire. They also suggested that the use of open-end questions should be kept to a minimum in self-completion questionnaires (McColl *et al.*, 2001).

Question sequencing

The context in which the question is posed does have an effect on the response (McColl *et al.*, 2001). Numerous recommendations are available in relation to questions order when designing a new questionnaire. If a pre-designed questionnaire is being used, the researcher is advised to use the same outline and frame.

Questionnaire appearance

The questionnaire appearance plays an important role in enhancing the response rate, especially when the questionnaire is self-completed and designed for visually-impaired participants. Appearance includes response format; layout and pagination; paper colour and quality; and length of the questionnaire. Small font and crowded pages will discourage the respondent from completing the questionnaires. McColl *et al.* (2001) listed a number of recommendations to improve the quality:

1. The cover design must present the title of the research and the identity of the organisation conducting it
2. The font should lies between 10 -14 points, but larger fonts could be used for visually-impaired respondents

3. The format of the questionnaire should be maintained consistently throughout

4. The whole question has to appear on one page

5. Similarly, linked questions are better kept on one page

6. Enough space should be provided in open-ended questions

7. Options (answers) for rating questionnaires should lie horizontally, and options for other closed-ended questions are better sorted vertically.

Covering letter

This letter is a very important part of the questionnaire, especially when this is administered by mail. The design and careful wording of this letter plays an important role in promoting respondents to take part in the study and provide as accurate and honest answers as possible (McColl *et al.*, 2000).

Piloting

Careful piloting of the questionnaire is a crucial and an indispensable stage during design, especially if this is a questionnaire that has not been used before. It is also equally important to pilot an existing questionnaire when this is being used in a different environment, e.g. a different population of respondents. Issues of interest during piloting include the clarity of the questions, participants' reaction to specific questions, and time needed to complete the questionnaire (Boynton, 2004).

2.4.3.3. Classification of questionnaires

Health questionnaires in general are classified by their target as generic or specific (Guyatt *et al.*, 1993; Fitzpatrick *et al.*, 1998; Garratt *et al.*, 2002). These are described in detail below. Generic instruments include health profiles and utility measures, while specific instruments are concerned with, for example, particular QOL dimensions, diseases, sites (in the body), or they may be individualised.

1. Health profiles

Generic measures are used to measure illness regardless of the cause or site. They usually measure several health domains presented in the form of multiple-item questionnaires such as the Short Form-36 (SF-36) Health Survey Questionnaire (Ware and Sherbourne, 1992).

2. Utility measures

Other types of generic measures have been developed for economic purposes. These tools are not designed to measure quality of life per se, but to evaluate the improvement to health-related quality of life gained by a particular intervention. Most of these tools produce a single index which is used by health policy makers to compare and justify different interventions based on the number of units of quality of life these treatments can provide (Bowling, 2001). An example of these tools is the Quality-Adjusted Life Years (QALYs) that was proposed in 1976 cited by Nord (1992). In the U.K., the corresponding measure is the Rosser Index of Disability (Rosser and Watts, 1972). The Euroqol (Williams, 1990) is also gaining popularity. It has been used recently to enable the estimation of health state values based on other health profile questionnaires such as the Oral Health Impact profile (Brennan and Spencer, 2006).

3. Dimension specific measures

These are concerned with particular aspects of health such as psychological well-being. They are usually scales and produce a single score. An example of these measures is the Beck Depression Inventory (Beck *et al.*, 1961). While pain is not considered a dimension on its own, some authors argue that it is, and thus classify pain specific measures in this category (Fitzpatrick *et al.*, 1998). An example of these instruments is the McGill Pain Questionnaire (MPQ) (Melzack, 1975).

4. Disease (condition or population) specific measures

As the name suggests, these target particular disorders or diseases and may explore all health-related effects, for example The Arthritis Impact Measurement Scales (AIMS) (Meenan *et al.*, 1980).

5. Site specific measures

These are designed exclusively for certain areas of the body, or for a certain medical speciality. Examples are The Oxford Hip Score (Dawson *et al.*, 1996) and the Oral Health Impact Profile OHIP (Slade and Spencer, 1994).

6. Individualised measures

These tools allow respondents to nominate aspects of their own life, which they consider to be the most important, and to weight the importance of these aspects. They have been developed in response to the argument that psychometric testing of scales developed at the expense of the relevance of the scales to the individual (Bowling, 2005). An example of these instruments is the Schedule for the Evaluation of Individual Quality of Life (SEIQOL) (Browne *et al.*, 1997). Some authors argue, however, that the hypothesis that these instruments have better psychometric properties than the standardised measure is unproven (Fitzpatrick, 1999).

2.4.3.4. Uses for health measurement questionnaires

Questionnaires can be classified by the purpose of the study as predictive, discriminative, or evaluative instruments (Kirshner and Guyatt, 1985).

1. Predictive tools

This type of instrument is used to "classify individuals into a set of predefined measurement categories when a gold standard is available" (Kirshner and Guyatt, 1985). These indices are often used as a diagnostic or screening tool to "predict" which specific individuals have or

will develop the condition or outcome of interest (Kirshner and Guyatt, 1985). These tools are employed to replace the gold standard measure which might be costly, difficult to use, or cannot be used in the early course of a disease. For example, the gold standard method to evaluate tooth loss is to examine the mouth. Alternatively, the predictive instrument might be asking the participants to report missing teeth on a questionnaire form (Worthington, 1998). In quality of life, however, there are no gold standards (Guyatt *et al.*, 1993). Therefore, the focus in the remainder of this section is on the other types of questionnaires, the discriminative and evaluative.

2. Discriminative tools

These are used to distinguish between individuals or groups on an underlying dimension; for example, to identify sufferers of a given disease from the normal population. Their main use is in cross-sectional surveys, with the aim of quantifying the illness burden across different groups. An example of these instruments is the short form OHIP-14 (Slade, 1997b).

Discriminatory measures do not contain items with high prevalence in the general population, since these items are common and may compromise the instrument's ability to distinguish between groups with severe and less severe impacts (Juniper *et al.*, 1996).

3. Evaluative instruments

These instruments are designed to measure longitudinal change in the condition of interest, and can be therefore used to assess treatment outcomes (Rogers *et al.*, 1998; Allen and McMillan, 2003). These tools are often used in clinical trials and cost-utility research (Slade, 1998).

Evaluative tools should be able to reflect the specific goals of the intervention of interest (Guyatt *et al.*, 1986). Moreover, they should contain items reported as the most important to the respondents (e.g. patients) (Juniper *et al.*, 1996). In this context, items with high prevalence in the population should be included, which allow for small changes in the condition to be reported (Locker and Allen, 2002). On the other hand, items used in the

discriminative tools, reflecting severe impact and low prevalence in the population are less amenable to change with intervention. These should not be used to evaluate treatment outcomes. For example, a question about pain severity asked to patients with orofacial pain will fail to discriminate between those with toothache, sore gum or mouth, or pain in the TMJ. However, this question can be successfully used to evaluate changes in pain severity when posed to the same patient on different occasions.

It is clear that the instrument's design and psychometric properties vary according to its purpose. These are summarised in Table 2.7 and explained in more details in section 2.4.5. The use of an instrument for a purpose other than what it was designed for might well compromise the results (Worthington, 1998; Locker and Allen, 2002). If a previously established instrument is to be used in a context other than that it was originally designed for, it is crucial that the instrument's properties are re-examined in the new context (Locker and Allen, 2002). The same applies when attempting to distil a lengthy instrument (Guyatt *et al.*, 1993), as discussed later.

Table 2.7. Major considerations in index construction and validation*

	Discriminative criteria	Predictive criteria	Evaluative criteria
Item selection	—tap important components of the domain —universal applicability to respondents —stability over time	—statistical association with criterion measure	—tap areas related to change in health status —responsiveness to clinically significant change
Item scaling	—short response sets which facilitate uniform interpretation	—response sets which maximize correlations with the criterion measure	—response sets with sufficient gradations to register change
Item reduction	—internal scaling or consistency —comprehensiveness and reduction of random error vs respondent burden	—power to predict vs respondent burden	—responsiveness vs respondent burden
Reliability	large and stable intersubject variation: correlation between replicate measures	—stable inter and intrasubject variation: chance corrected agreement between replicate measures	—stable intrasubject variation: insignificant variation between replicate measures
Validity	—cross-sectional construct validity: relationship between index and external measures at a single point in time	—criterion validity: agreement with criterion measure	—longitudinal construct validity: relationship between changes in index and external measures over time
Responsiveness	—not relevant	—not relevant	—power of the test to detect a clinically important difference

*** From (Kirshner and Guyatt, 1985)**

Different formats of questionnaires are proposed to measure changes over time, but there is no universal consensus over what method to use (Locker, 1998; Slade, 1998; Allen, 2003). The various types are:

1. **Comparisons of "before" and "after".** This is the simplest format, where group summary scores are calculated for the study participants on two or more occasions. The drawback of this approach is that it treats the scores before and after the event of interest as independent, and does not allow for "within-subject" changes to be calculated. Therefore, in a given intervention, responders and non-responders cancel each others' scores.
2. **Computing individual change scores.** This is done by subtracting the baseline score from the follow-up score for each individual, allowing for identification of those whose scores improved or lowered.
3. **Global transition judgment.** This is the patient's (respondent's) retrospective assessment of overall changes in their condition during the study period. This method is considered better than the previous as it allows the respondent to decide their own reference points, i.e. before and after the event of interest (Slade, 2002). It is criticised, however, for being a retrospective assessment that is prone to bias (see section 2.5.1) and which does not take into account fluctuation of the condition over time (Dolan *et al.*, 1998).
4. **Retrospective judgement of extent of change.** This is an extension of the previous method, where respondents rate the extent of change numerically rather than just report its occurrence or direction.
5. **Categoric enumeration of change** (Slade, 1998). In this method, within-subject changes are considered as in method two. The author argues that domains of quality of life do not necessarily change in the same direction. Therefore changes may cancel each other even within the same individual. Slade (1998) exemplifies this by a person who lost some teeth which reduced pain but introduced difficulty in chewing. He recommends categoric enumeration of change by considering both increment and decrement in scores.

2.4.3.5. Shortening lengthy questionnaires

Lengthy questionnaires are deemed to be too costly in large national surveys, inappropriate for compromised or aged participants, subject to large item non-response in cohorts with specific disorders, and difficult to use in clinical practice (Locker and Allen, 2002). For these reasons, shorter versions of composite measurement scales have been developed either to replace long versions, or to make them more specific for certain disorders or populations.

Shorter versions designed as condition-specific measures have better responsiveness than the original long questionnaire (Allen and Locker, 2002; Allen, 2003). Also, the administration of short versions of the questionnaire is easier. While this is a benefit, it can be argued that reducing the number of items may adversely affect measurement. Obviously, the content validity of the short-form measures is compromised (Locker and Allen, 2002). In addition, the reliability of an existing questionnaire tends to decrease as items are eliminated (Nunnally and Berbstein, 1994).

Various approaches have been developed to produce shorter versions of composite measurement scales while maintaining psychometric characteristics. Different methods, however, maintain different psychometric properties. The decision about which approach to use depends on the purpose and required properties of the short-form version (Guyatt *et al.*, 1993; Juniper *et al.*, 1997; Locker and Allen, 2002).

For discriminative instruments, the aim is to produce a shorter version which is able to classify respondents (patients) in a similar way to the original. Therefore, the researcher should examine the extent to which variance in scores in the full instrument is explained or predicted by scores of the short-from (Guyatt *et al.*, 1993). If the intention, however, is to produce an evaluative short version, psychometric properties of concern should be the instrument's validity in the new context and its responsiveness to change with treatment in comparison to those properties of the full version (Guyatt *et al.*, 1993).

The literature reports three popular approaches for item reduction: the *Item-Impact* analysis (Juniper *et al.*, 1997), the *Regression Analysis* and the *Factor Analysis* (Coste *et al.*, 1997). The three methods and their favoured applications are described below.

The item-impact analysis (Juniper *et al.*, 1997) identifies items that are most important to the respondents (Locker and Allen, 2002). In this method, respondents (e.g. patients) are given a pool of statements (in this case the long form instrument) and are asked to indicate the items which they have experienced over a given period of time, and to rate these items on a Likert-type scale of importance. The item-impact score for each item is calculated by multiplying the proportion of respondents experiencing this item by its mean importance rating (Juniper *et al.*, 1997). As an example, this method was used to derive a population specific short form of the 49-item Oral Health Impact Profile (OHIP) (Allen and Locker, 2002).

Regression analysis is a "method for describing the relationship between two or more variables. It estimates the impact of one variable on another" (De Vaus, 2001). In other words, it evaluates how much of the variance in a variable is explained by the variance of another. This method has been used to shorten long questionnaires by identifying individual items that contribute the most to the full questionnaire. In this context, items which explain most of the variance in the total questionnaire are considered important, and those which explain little variance are considered redundant and are eliminated. For example, Slade (1997b) used this method to shorten OHIP-49 to OHIP-14.

Factor analysis is a "complex method of reducing a large set of variables to a smaller set of underlying variables referred to as factors" (De Vaus, 2001). This analysis can be conducted in a number of ways, but they all serve the same purpose, and that is to identify the underlying factors of a large set of variables, e.g. questionnaire items. For example, this method has been used to identify the possible dimensions of oral health-related quality of life

(OHrQOL) (John *et al.*, 2004a). The authors identified four dimensions (factors) and derived a 21-item short version of the original OHIP-49.

In a review of the methodology of shortening composite measurement scales (Coste *et al.*, 1997), the authors emphasised that the regression and factor analyses are most suitable when the original scale is the gold standard for measuring the phenomenon of interest. Here, it is essential to optimise the "criterion" validity of the short-form by ensuring a high degree of correlation between the shortened and original form. In contrast, when the original instrument is not the gold standard, these analyses are inappropriate and other methods are preferred.

If the short-form questionnaire is evaluative, i.e. is to be used as a treatment outcome measure, it is essential that this is based on what patients feel is important (Guyatt *et al.*, 1986; Juniper *et al.*, 1996). For this reason, the item-impact method is preferred (Locker and Allen, 2002) as it is based on patients' ranking of items according to their importance (Juniper *et al.*, 1997). Clearly, evaluative instruments should employ items with distinguished responsiveness to change with the intervention of interest (the *Effect Size*), and which have low floor effects (see measurement properties - section 2.4.5.3).

To this end, the performance of two short versions of OHIP has been examined (Allen and Locker, 2002; Locker and Allen, 2002); one derived using the regression analysis [OHIP-14 (Slade, 1997b)], and another obtained using the item-impact method [OHIP-Edent (Allen and Locker, 2002)]. It has been concluded that the item-impact method yielded an instrument that is better in detecting change with treatment, and that has lower floor effects.

How short is short?

There are no clear guides to how much one can cut down an instrument. It is important, though, to maintain the same framework of the original form (Juniper *et al.*, 1997; Slade,

1997b; Locker and Allen, 2002). As mentioned before, the shorter the instrument, the weaker the psychometric features.

For health instruments, it has been suggested that each of the index's dimensions needs to be represented by three to four items (Juniper *et al.*, 1997). It is also argued that increasing the number of items per domain decreases the risk of floor effects (Bindman *et al.*, 1990). On the other hand, reducing the instrument to two items per domain would maintain responsiveness and reliability of composite measurement scales at an acceptable level, with a trade-off of reduced construct validity and increase in sample size requirements (Moran *et al.*, 2001). This argument is supported by the fact that some instruments were shortened to two items per dimension and were still valid and widely used, e.g. OHIP-14 (Slade, 1997b).

2.4.4. Scales

While many self-administered instruments, e.g. questionnaires, are composed in general of a battery of indicators, scales are a more sophisticated and perhaps more informative self-report method. Scaling is the assignation of numbers to responses for given questions (Bowling and Ebrahim, 2005). This is essential to enable qualitative data to be analysed in a mathematical or statistical way.

Scales can be single or multiple questions posed to respondents on certain attitudes, feelings and perceptions. They can be uni- or multidimensional in relation to the number of domains they cover. An example of unidimensional scale is a single-item scale of pain intensity (see section 2.4.4.4). Multidimensional scales are usually in the form of multiple-item questionnaires. Examples of these are presented in sections 2.4.3 and 2.4.4.4). Some single-item scales can be used, however, as an alternative for long multidimensional questionnaires. These are usually referred to as "*Global Self-Rating*".

In global scales subjects respond to a comprehensive question about their perception of the phenomenon being investigated. Using this method it is assumed that the respondents will interpret the dimensions of QOL that are important to them, make judgement of the impact of their condition on these dimensions, and incorporate this judgement into a single score (Slade, 2002). The validity and reliability of this approach has been addressed in a large number of studies on health in general as well as on health related quality of life (Cunny and Perri, 1991; de Boer *et al.*, 2004).

Global scales are very popular in health and its related quality of life research, mainly because they are short, simple, and easy to administer and to analyse. Other advantages include permitting respondents to assimilate multidimensional experiences and choose those which are most prominent (Slade, 2002). Furthermore, they allow for positive ranking of

health, as opposed to measuring the extent of illness, something that fits with ICF the International Classification of Functioning, Disability and Health mentioned earlier (World Health Organization, 2006). It should be noted, however, that detailed multiple-item scales are inevitable to obtain a clearer picture of the individual or QOL, which can be masked by overall rating. For example, in a study on the impact of dental and other oral disorders (Slade, 1998), it has been found that while tooth loss may affect chewing, it decreases pain. Reporting both improvement and deterioration of health simultaneously is only possible with a detailed multiple-item questionnaire, and was not possible with a summary single-item scale. It is therefore recommended that multiple-item instruments should be used whenever possible (Cunny and Perri, 1991).

2.4.4.1. Scales' data

Whether it is a uni- or multidimensional, single- or multiple-item scale, scales' data can be categorised in four levels: nominal, ordinal, interval, and ratio (Bowling, 2002; Bowling and Ebrahim, 2005).

1. Nominal and categorical scales

These scales classify responses into groups and use numbers to distinguish the different groups but do not give them any order. A simple example of nominal classification is grouping patients by their gender where male patients are coded by 1 and females by 2. These numbers do not imply that females are greater or lesser than males. They simply denote that patients who have the same number are from the same gender. Data can be dichotomous such as in yes/no questions or descriptive (race, eye colour, etc.).

2. Ordinal scales

As the name suggests, these scales order responses in a hierarchy. Items in this type of scales are ranked with respect to their importance, severity or frequency from least to most. For

example, reported pain can be mild=1, moderate=2 or severe=3. There are no definite intervals among the groups, i.e. it cannot be concluded that the difference between mild pain and moderate pain is identical to that between the latter and severe pain. Ordinal scales such as *Likert-type scales,* are adopted in a large number of HRQoL instruments. Yet, the optimal scoring interval is controversial. A good scale should provide the respondent with an adequate number of choices. On the one hand, limited choices such as "Do you agree or disagree" can lead to confusion and error in the measurement. On the other hand, redundant options might result in meaningless responses. Most of HRQoL instruments offer four to seven options.

3. Interval scales

These assign definite intervals among the ordered responses. In other words, numerically equal distances on the scale represent equal distances on the attribute of interest. The Visual Analogue Scale VAS which is used to measure pain is an example. Patients are asked to indicate their pain level on a line of 100 mm length. A score difference of 10 is the same whether it is between 60 and 70 or 80 and 90. The absolute value of the attribute, however, is either unknown, meaningless, or given arbitrarily.

4. Ratio scales

These are the highest level of scales, and the most accurate (Bowling, 2002). They are suitable for attributes that have absolute values, for example, the patient's weight. Unlike the interval scales, the zero point here is the original value. Thus, a value of zero suggests here that the respondent has none of the measured attribute.

2.4.4.2. Statistical analyses of scales

From a statistical point of view, analytical options rise with increasing levels of scales. To start with, only simple statistics can be performed for nominal data including cross-tabulation and frequencies. Non-parametric procedures can be considered at the ordinal level, but these

data can only be ranked, not sorted, and cannot be averaged or mathematically manipulated (Bowling, 2002). The median is, therefore, the most appropriate statistics to describe the central tendency of scores. Optimally, interval and ratio scales can be analysed by parametric statistics (Bowling, 2002). Any statistical measure can be used to analyze data from ratio level.

However, no rating scales achieve ratio scale level (Bowling, 2002). Ideally, for accuracy and statistical considerations, measures should be developed using interval or ratio scales. Yet, obviously, this is difficult in social sciences. But there are still numerous techniques available to help in transforming nominal data into interval and ratio levels (Bowling, 2002).

2.4.4.3. Scoring scales

The simplest way to get data from a scale is by summing its items' scores. Some items, however, may be of more importance than others and should therefore contribute more to the total score. Attributing equal scores to the scale items would lead to assigning the same weight to items that might not be of the same importance to the patient. A given score can be arrived at in different ways. For instance, someone who is physically disabled, but who has a joyful character may score the same as someone else who is mobile but suffers a psychological problem. Though this may be meaningful when assessing workload, it has no value in assessing the individual's status. In the field of temporomandibular disorders an index has been developed to measure the severity of the condition and to monitor treatment outcomes: The Temporomandibular Index (Pehling *et al.*, 2002). While this index has some merits, it suffers by trying to produce an overall score for the condition. This overall score does not provide any meaningful result, as the patient can reach the same score whether he/she suffers a disc displacement or a myofascial pain.

This problem can be avoided by assigning different weights to the items according to their importance to the patient. For example, three oral health-related quality of life measures, the Dental Impact on Daily Living (DIDL) (Leao and Sheiham, 1996), the Oral Health Impact

Profile (OHIP) (Slade and Spencer, 1994) and the UK oral health-related quality of life [OHQoL-UK (W)] (McGrath and Bedi, 2001), consist of items with weights. Weighting is often based on Thurstone's method of paired comparisons (Allen and Locker, 1997; Slade, 1997a).

On the other hand, it is argued that weighting brings some difficulties and complexities, and may not be feasible in clinical settings (Streiner and Norman, 1995). These authors suggest that weighting might be useful for indices with fewer than 40 items, but those with more items may not benefit from this additional procedure. Their argument is supported by some research on item weighting in OHrQOL measures (Leao and Sheiham, 1996; Allen *et al.*, 2001; Aaron *et al.*, 2004; McGrath and Bedi, 2004).

2.4.4.4. Examples of scales used for the measurement of pain

Single-item scales

Various types of single-item pain scales have been developed (Turk and Melzack, 2001) including:

1. The Verbal Rating Scale (VRS)

This scale provides a choice of several options, preferably 5-7 (Bowling, 2002) from which the patient is asked to select the one that best describes the intensity of their pain (e.g. none, mild, moderate, severe, very severe) or its frequency (e.g. never, hardly ever, occasionally, often, very often). This type of scales is often referred to as Likert type (Bowling, 2002).

2. The Numerical Rating Scale (NRS)

This scale consists of a sequence of numbers (e.g. 0 – 10 or 10 – 100) form which the patient is asked to circle a choice that corresponds to the degree of suffering.

3. The Visual Analogue Scale (VAS)

This is a line, preferably horizontal and 100 mm long (Bowling, 2002), that is anchored at its

ends by the extreme limits of the phenomenon being measured, for instance the level of suffering *"no pain"* and *"the worst imaginable pain"*. Respondents are asked to mark on this line the extent of their suffering (e.g. intensity or unpleasantness of pain). The scale score is represented by the distance between the start point and the indicated mark. It has been found that lines shorter than 10 cm tend to produce greater error variance and that intermediate points should not be defined to avoid clustering of scales.

While the three methods have well established validity and reliability, VAS is the favoured instrument (Turk and Melzack, 2001). This is due in part to the observation that NRS leads to response clustering. In other words, respondents are derived to choose one option (one number) as opposed to giving their exact measure of the phenomenon as the case in VAS. By this, NRS imitates Likert-type scales (Wewers and Lowe, 1990). Another problem is patient ability to remember previous answers thus leading to a bias (Gendreau *et al.*, 2003). Also, the verbal descriptions used in the VRS may be interpreted differently by different respondents, thus affecting the validity of the scale (Bowling, 2002).

The VAS technique, which became widely used in 1970s, has been adopted as an alternative for the ordinal scales and as a solution to the scoring debate. While some argue that VAS is an ordinal scale, others believe it also meets the criteria of interval and ratio scales (Maxwell, 1978; Wewers and Lowe, 1990).

VAS is often used for assessing clinical symptoms, especially pain (Olsen *et al.*, 1992). It has been also used for the evaluation of sleep, functional performance, behaviour and affective state, e.g. mood and anxiety. This technique better mirrors the continuous nature of feelings and attitudes, has been extensively tested for validity and sensitivity to change with treatment in research on feelings (Zealley and Aitken, 1969).

It is not surprising that VASs have been used in a large number of TMD research, including measurement of pain intensity (Dao *et al.*, 1994; Wassell *et al.*, 2004) and perceived improvement with treatment (Dahlstrom *et al.*, 2000; de Boer *et al.*, 2004).

Multiple-item scales

McGill Pain Questionnaire (MPQ) (Melzack, 1975) is a well established and widely used measure of pain. The author of this instrument argues that assessing pain merely in a numerical way is limited and that the nature of pain is an important factor in the experience of pain and has to be considered in the evaluation of pain. To measure this nature, MPQ enables the respondent to choose from an extensive set of verbal descriptors of pain. These are then transformed into numerical codes in order to yield a score. MPQ does not measure pain intensity and nature only, but also pain-related disability (LeResche, 1997a). Both MPQ and its short form (Melzack, 1987) have been used to measure pain and related disability in a number of orofacial pain conditions including TMD (Rudy *et al.*, 2001; Van Wijk and Hoogstraten, 2002).

2.4.5. Psychometric characteristics

Researchers agree that self-report instruments have to meet certain criteria before they can be used in research. Most importantly they have to be valid and reliable, but other measures are also important.

2.4.5.1. Validity

In simple terms, a valid instrument is one that measures what it is supposed to be measuring. Validity is not a single feature. It has several aspects that are listed below (Litwin, 1995; McColl *et al.*, 2000; Fayers and Hays, 2004)

1. Face validity

The questions asked here is "whether on the face of it" the instrument is measuring what it is supposed to measure. This is only a primary step and is usually examined by asking a group of ordinary non-specialised people to check if the instrument items appear appropriate.

2. Content validity

This type of validity explores if the content of the instrument, e.g. the questionnaire items, covers everything it should, and does not include irrelevant elements. This is assessed by a panel of informed people including specialists and members of the target population, e.g. patients.

3. Construct validity

This is the most rigorous approach to establishing validity (Guyatt *et al.*, 1993), which investigates if the results obtained by the instrument confirm expectation. Two methods are used to assess this aspect of validity.

In the first method the results of the target group are compared with another group that is expected to produce different results e.g. patients against healthy controls. This method is known as *Known-group validity*, *group validity* or *discriminant validity* and it is suitable to validate instruments with discriminative purpose (Guyatt *et al.*, 1993).

The second method, *multi-trait multi-method analysis* (Campbell and Fiske, 1959) involves examining the behaviour of the investigated instrument against another simultaneously administered instrument of known validity (McColl *et al.*, 2000). High correlation is expected amongst instruments that are measuring the same phenomenon or conception (*Convergent Validity*). Weak correlation is expected between instruments measuring dissimilar concepts (*Divergent Validity*). Different methods are available to measure the correlation coefficients depending on the data type. Pearson's correlation coefficient "*r*" is appropriate when at least one of the variables is at the interval (continuous, numerical) level, while Spearman's rho correlation coefficient "r_s" is the test of choice for ordinal data.

To validate an evaluative instrument, changes over time detected by this tool should correlate to those measured by another instrument of established properties, given that both instruments are measuring the same phenomenon (Guyatt *et al.*, 1993; Worthington, 1998).

The correlation between variables (e.g. instruments) using either Pearson's or Spearman's correlation coefficients varies between -1 and +1, where the sign indicates the direction of correlation, i.e. positive or negative. The values between 0 and the absolute 1 and their meanings are presented in Table 2.8 (Riffenburgh, 2006).

Table 2.8. Values of Pearson and Spearman's correlation and their interpretation

0			No relationship
0.10 to 0.29	or	-0.10 to -0.29	Small
0.30 to 0.49	or	-0.30 to -0.49	Medium
0.50 to 0.99	or	-0.50 to -0.99	Large
1	or	-1	Perfect

4. Criterion validity

This type of validity is considered as a special case of construct validity (Fayers and Hays, 2004). It examines the inter-correlation between the investigated instrument and a gold

standard measurement of the same concept applied concurrently on the same group. An example of this is testing a self-reported statement of high temperature against the measurement on a thermometer. HrQoL suffers, however, from a lack of gold standard measures. Hence, the assessment of the validity of HrQoL measures is mainly limited to content and construct validity (Fayers and Hays, 2004).

Nevertheless, when testing the validity of a new shorter instrument that has been derived from an original one, the "mother" version is considered the gold standard (Coste *et al.*, 1997).

2.4.5.2. Reliability

A reliable instrument is one that measures the concept of interest in a consistent or reproducible manner. As with validity, there are different aspects of reliability (McColl *et al.*, 2000; Fayers and Hays, 2004):

1. Stability or test-retest reliability

This analysis explores the reproducibility of the obtained results when the instrument is administered to the same person on two occasions within which the condition being investigated was "relatively" stable.

There has been considerable debate in the literature regarding the most appropriate choice of the reliability coefficient (Streiner and Norman, 1995). One method to assess this reliability is to calculate the correlation coefficient CC (or Pearson's coefficient) between the two stages' data (Worthington, 1998). In essence, the correlation test examines the ratio of between-subject variance to total variance, which includes both between- and within-subject variance (Guyatt *et al.*, 1992). The reliability, denoted "*r*", is considered good if *r* values ≥ 0.70 (Litwin, 1995). Streiner and Norman (1995) argue that Intraclass Correlation Coefficient, ICC is a better measure of reliability than Pearson's *r* both technically and mathematically.

While correlation coefficients measure the strength of the association between the two variables, another analysis is needed that determines whether the change in scores between the test and retest is statistically significant. A paired *t*-test can be used for numerical normally distributed data, while Wilcoxon is the nonparametric equivalent test for ordinal data (Pallant, 2005).

Testing the stability of a scale is challenging in the clinical settings because the interval between the two occasions has to be long enough to avoid memory effect, i.e. patients recalling their initial answers, and not too long so the condition does not change.

The test-retest reliability for established instruments must be re-examined when they are used in a different population than that they were originally tested on (Litwin, 1995).

2. Inter-rater reliability and intra-rater reliability

These are special cases of the test-retest reliability. Inter-rater reliability is also known as *inter-observer, between-observer, or between-interviewer reliability* and refers to whether different examiners will reach the same assessment for the same observation (e.g. patient) given no changes in the condition. Inter-rater reliability for agreement on ratings is measured by *Cohen's Kappa* (Cohen, 1988; Riffenburgh, 2006), which evaluates agreement in excess of that is expected to occur by chance. Kappa of 0 implies no agreement better than chance, while 1 implies perfect agreement. Other values and their interpretation are given in Table 2.9 (Landis and Koch, 1977; Cohen, 1988). Different formulae to calculate Kappa have been developed to allow the assessment of agreement between two or more assessors; for two or more categories (Riffenburgh, 2006).

Table 2.9. Values of Cohen's Kappa and their interpretation

0	No agreement
< 0.2	Poor
> 0.2 ≥0.4	Fair
> 0.4 ≥ 0.6	Moderate
> 0.6 ≥ 0.8	Good
> 0.8 ≥ 1	Very good
1	Perfect agreement

Intra-rater reliability, also known as *intra-observer, within-observer, or within-interviewer reliability* refers to whether the observer can obtain the same results from the same participant on different occasions, given nothing changed in the meantime. This is mostly required for predictive instruments when the purpose is to classify an observation into the same category.

The Intra-rater reliability is assessed by a statistic known as *Cronbach's coefficient alpha "α"* (Worthington, 1998), named after the psychometrician who first reported it in 1951. Alpha lies between 0 and 1, but should be above 0.70 for the instrument to be reliable (Nunnally and Berbstein, 1994; Streiner and Norman, 1995).

3. Internal consistency

This indicator, sometimes named *Homogeneity*, reflects the harmony between the instrument's elements. In other words, it investigates if all the questions of a questionnaire are measuring the same or a related phenomenon, thus achieving results that are inter-correlated. This consistency is also assessed by *Cronbach's α* (Litwin, 1995; Riffenburgh, 2006). If the instrument's α is low, it can be improved by adding more items or re-examining the existing items for clarity.

Similar to the test-retest reliability, it is essential to measure the internal consistency for each new instrument, as well as for existing instruments with established internal consistency when used in a population other than that for which they were designed (Litwin, 1995).

2.4.5.3. Responsiveness

1. Sensitivity to change

This feature is particularly relevant to evaluative instruments. In longitudinal research, it is fundamental that the instrument is capable of detecting change (both improvement and deterioration) in the condition of interest (Fitzpatrick *et al.*, 1992b). Similar to the test-retest reliability, when measuring sensitivity, variance due to real change in the condition has to be distinguished from that related to measuring errors (Guyatt *et al.*, 1992). To this end, it is not only important to examine the significance of score change pre- and post treatment, but it is essential to determine the true amount of change captured by the instrument. This is measured by calculating the *Effect Size*, which results from dividing the mean of score change by the standard deviation of the post-intervention score (Fitzpatrick *et al.*, 1992b; Allen and Locker, 2002). The effect size is considered small, moderate and large if it measures respectively ≤ 0.2, 0.3 to 0.6, and ≥ 0.7 (Cohen, 1988; Allen and Locker, 2002). Another way to examine the sensitivity to change of larger multiple-item instruments is to administer a parallel single-item global scale of self-reported change (Dolan *et al.*, 1998; Locker, 1998).

2. Ceiling or floor effects

The responsiveness of an instrument can be compromised by ceiling or floor effects (Guyatt *et al.*, 1993; Locker and Allen, 2002). That is, patients may report the highest possible scores on a measurement of function, while their quality of life is impaired by other factors that the instrument failed to show – a ceiling effect. On the other hand, patients reporting lowest possible scores in an instrument may feel worse than others, or may get worse with time, yet the scale fails to demonstrate this. This is called the floor phenomenon (or effect) (Bindman *et al.*, 1990).

It is crucial to employ questions that are responsive to the condition being investigated, and to use measurement units that allow the detection of small but clinically important changes.

2.4.6. Other methodological considerations

The researcher has to be aware of other research and ethical issues that are of importance.

2.4.6.1. Confidentiality

Whether in-depth interviews or questionnaires are involved, great care should be exercised during data collection and handling. These data often comprise personal information and opinions and should be treated as confidential. Data forms should always be kept separate from any information identifying participants. Only codes can be attached to data forms, which can be used when necessary to trace back the respondent. Codes of practice should be followed both from national sources, e.g. data protection legislation (The Information Commissioner's Office in the UK, 1998); and internal sources, e.g. relevant University and Hospital Trust guidelines.

2.4.6.2. Administering and data handling errors

Various types of errors can occur during administration of questionnaires and data processing including coding and data entering. Every effort should be made to avoid or minimise mistakes. It is noteworthy that computer assisted interviews or questionnaires have major advantages over traditional methods in this context.

2.4.6.3. Bias

Bias can occur at any point of research, especially in the sampling stage. Bias can still occur even when both the instrument and the sampling method are valid. An example of this is the *Auspices Bias* seen for instance when patients tend to report higher satisfaction with treatment. This is especially the case in interviewer-administered questionnaires. The interviewer has to be perceived by the interviewee as neutral with regard to the subject matter of the interview. In this context hospital staff are advised not to interview their own patients regarding satisfaction of care for example, as this might introduce bias by patients attempting to please their hospital staff (Fitzpatrick and Boulton, 1994; Britten, 1995). Similarly, bias can occur when respondents may under-report socially undesirable behaviour in what is called *Social*

Desirability Bias. *Recall Bias* is another type of bias and is observed with questions involving memory (McColl *et al.*, 2000).

2.4.6.4. Sample size

The sample of the research participants should be designed at the outset so that the emerging results can be generalised with confidence. One aspect of this design is the number of participants in a study that is required to obtain precise and reliable data. This is often controversial. On the one hand, the bigger the number the lesser the standard error. On the other hand, increasing the sample size consequently increases costs for data collection, processing and analyses.

Numerous factors control the suitable sample size (Fink, 1995; Bowling and Ebrahim, 2005). These include the following:

1. **Objectives of the study:** the more objectives the study aims to achieve, the more variables introduced, with the consequent need for sufficient data.
2. **Number of subgroups of participants:** it is important to ensure fair representation for all groups in the study.
3. **Type of data:** data analysis relies grossly on its type, i.e. qualitative or quantitative. While a small number of participants is adequate to generate large amounts of qualitative data, statistical analyses of quantitative data often require a greater amount of data to be valid.
4. **Non-response rate:** If this is expected to be considerable, over-sampling is often needed to obtain the desired response. This is often the case with postal questionnaires which have a lower response rate than "in-house" questionnaires, and longitudinal studies where participants fail to complete the study.
5. **Administration mode:** face-to-face questionnaires are costly in comparison to the self-completion ones, and any additional participant will attribute markedly to the total cost of the study.

Estimating sample size

The size of the sample should be calculated at the design stage of the study using a power calculation. The *test power* is its probability to detect differences between the investigated groups at a given level of significance (Petrie *et al.*, 2002). This power should exceed 80% (Crichton, 1993; Petrie *et al.*, 2002). The statistical significance of the study results relies also on the true difference between the populations being investigated and the level of significance. The level of significance is often selected at 0.05 (Petrie *et al.*, 2002).

Having determined all the above variables, calculating the desired sample size can be performed using Altman's nomogram (Altman, 1991). Alternatively, some statistical software packages are available to do this calculation.

2.4.6.5. Missing data and enhancing response rates

All self-report measures suffer from low response rates to some degree (McColl *et al.*, 2000). This is particularly seen with postal and follow-up surveys where completing the questionnaire or the diary relies completely on the respondents' cooperation. The response rate may be acceptable sometimes, or considered unacceptable in others. This is when it is thought that the non-response rate can affect the validity of the results, while the questionnaire itself is a valid instrument.

There are two types of response rate: *Overall (questionnaire) response rate*, defined as the proportion of respondents to the total number of eligible respondents; and *Item response rate* representing the proportion of answered questions (items) to the total number of items (McColl *et al.*, 2000). Different approaches are recommended to improve the response rate (McColl *et al.*, 2000; Boynton, 2004):

1. Minimising the respondents' efforts to complete the survey and to return it, i.e. avoiding lengthy and complicated questions and providing pre-paid self-addressed envelopes

2. Maximising the respondent's rewards, for instance by emphasising the importance of the survey to the patient and all sufferers of the same problem, by personalising the covering letter, by expressing appreciation in all communications, and by informing the participants of the research results
3. Establishing trust between the researcher and the respondents by means of establishing the credentials of the researcher (e.g. using headed papers in correspondence, naming of researchers), and by providing endorsement by well-regarded organisation.

The latest technologies in obtaining and entering data help to reduce the amount of missing data. For example, computer-based questionnaires can be programmed to alert the respondent to uncompleted fields, or may even be programmed so they do not allow the respondent to proceed before completing all required fields.

Ultimately, it is not uncommon to have missing data and it is the researcher's responsibility to assess the volume of these data and how they affect the outcome of the study. If missing data lie within the acceptable criteria set by the research, a decision has to be made whether these data are to be excluded or replaced. Replacing missing data is known as *Data Imputation*. Different methods for data imputation are available in current statistical packages (Riffenburgh, 2006).

2.5. Assessing health in chronic disorders – timing of treatment outcome assessment

2.5.1. The review period

Establishing the time at which treatment outcomes are to be assessed plays a critical role in the reliability of the assessment. On the one hand, sufficient time must be allowed for the treatment to achieve the desired outcomes. On the other hand, the assessment should not be delayed too much, to enable the clinician to change treatment plans within acceptable time should the treatment prove unsuccessful.

The time until assessment, the so called *Review Period* must be based on evidence and not chosen arbitrarily or this might well affect the reliability of the assessment. In TMD, however, this period appears to have been derived from clinical experience and expert opinion, but with no supporting evidence. For example, to test the sensitivity to changes with treatment for a "new outcome measure for TMD" (Pehling *et al.*, 2002), the authors stated that "post-treatment testing was completed when the patient and treating doctor determined that a healing plateau had been achieved". Their review period averaged around six weeks after the start of treatment, but again without supporting scientific reference.

Two systematic review studies for randomized controlled trials and splint therapy in TMD (Forssell *et al.*, 1999; Al-Ani *et al.*, 2004) have found that follow-up periods varied between four and 52 weeks (in the reviewed articles alone). To the best of my knowledge there is no literature that attempts to identify optimal review periods for TMD treatments.

The optimal review period is principally dependent on the end point of treatment, ideally, the time when a cure has been achieved. However, in the case of chronic disorders, symptoms often linger and a cure may not be expected. In this scenario, clinicians will endeavour to minimize symptoms and their impact on patients' lives. Therefore, establishing the optimal

review period for chronic conditions should be based on the point when acceptable or important outcomes are achieved.

In this context, defining "important outcomes" becomes essential, not only to determine the optimal review period, but also to produce potential cut-off points to sort treatment responders from non-responders, a measure that is increasingly adopted in clinical trials (Farrar, 2000). Furthermore, it constitutes a standard outcome measure to compare different clinical trials (Farrar *et al.*, 2000; McQuay, 2005).

Different attempts have been reported in the literature to define the "Clinically Important Difference - CID" in pain treatment outcomes. Unfortunately, there is no consensus, however, over what CID means, and what volume of improvement should be considered as CID (Farrar, 2000; McQuay, 2005). Firstly, while some studies defined CID as the minimum acceptable level of change in symptoms that can be reported as improvement (Todd and Funk, 1996), others defined CID as the minimum desirable achievement from a particular treatment. They used this to draw a cut-off point between improvers and non-improvers (Moore *et al.*, 1997; Farrar *et al.*, 2000). Consequently, lacking a clear definition of CID lead to disagreement about its value. For some (Todd and Funk, 1996) it meant a minimum reduction of 18 out of 100 mm in the pain intensity as measured by the visual analogue scale. Other workers suggested that CID should be based on percentage of improvement (Goldsmith *et al.*, 1993). They concluded that "an improvement of at least 36% should be clinically important for patients" in rheumatoid arthritis clinical trials. A shortcoming in the methods in the above studies is reliance on the professionals' assessment of the patient's pain instead of being patient-based (Todd and Funk, 1996; Farrar, 2000).

The value of CID ranges in the literature between 25% and 75% pain reduction without clear rationale (Farrar, 2000). In particular, the 50% threshold has been adopted in different studies (Freedman *et al.*, 2003; Mason *et al.*, 2004a; Mason *et al.*, 2004b) yet again there is no justification.

Different approaches have been taken to establish CID. A meta-analysis study of analgesic clinical trials compared different patient-based outcome measures including pain intensity and pain relief. The authors suggested 50% reduction in pain intensity as a cut-off point to group patients as "responders" or "non-responders" (Moore *et al.*, 1997). The study also suggests that the number of responders compared to non-responders is an important outcome measure of clinical trials. Another approach has been used with complex regional pain syndrome type one. In this case patients' global perceptions of successful and unsuccessful treatment were related to a reduction in VAS pain scores (Forouzanfar *et al.*, 2003). The results show that a pain reduction of at least 50% and an absolute reduction of 3 cm on a 10 cm VAS are accurate in predicting successful treatment.

The concept of using "rescue medicine" has been used to establish the aforementioned cut-off point between improvers and non-improvers (Farrar *et al.*, 2000). Rescue medicine is an additional drug that patients can take if their routine pain management fails to reduce pain sufficiently. As such, it can act as a gold standard for the clinical efficacy of routine treatment. This test, which has been used for cancer treatment, produces dichotomous data with clear cut-off point (i.e. whether the treatment worked or failed in improving symptoms) against which cut-off points from other pain scales can be standardised (Farrar *et al.*, 2000; Farrar *et al.*, 2003).

To summarize, it is essential to establish the optimal time for treatment outcome assessment (review period). Arbitrary decisions about this time may have important effects on the reliability of the outcome assessment. Different methods have been suggested to establish the review period, however, there is no consensus. The concept of Clinically Important Difference, CID, still remains nebulous. Perhaps Pehling's idea (Pehling *et al.*, 2002) of delaying post treatment testing until a healing plateau had been achieved offers a good outcome. Nevertheless, as with CID, further work is needed to define this more clearly for TMD and its sub-diagnoses.

2.5.2. The reference period

Treatment outcome can be assessed by examining signs and symptoms at the time of the assessment, or by assessing the patient's experience of these over a past period of time. This is especially the case with a self-report instrument, such as a questionnaire, where patients will be asked to report their experience retrospectively with reference to a specific period, hence called the *Reference Period*. Determination of this period has to be done in conjunction with the outcome assessment timing. It too has to be evidence-based to provide consistency across different studies of the same condition.

The reference period should allow patients enough time to make a meaningful assessment of their experience. On the other hand, it should not be too long or it will include the time near the start of treatment when no or little improvement had been achieved. It is therefore essential that this period allows for a realistic CID to be reached and either maintained or exceeded for some time to enable the patient to make a valid assessment of the hopefully improved condition.

If a previously designed questionnaire is to be used, which is not specifically designed for a TMD population, the validity of its original reference period has to be established. Generalising the same reference period blindly to a population other than the original one may well compromise the results. The main questionnaire used in this project, OHIP, has been originally designed and used for patients with oral and dental problems in general with a reference period of 12 months (Slade and Spencer, 1994; Slade, 1997b). As part of validating the German version of OHIP, John *et al.* (2002) reported that a one month reference period had the highest internal consistency.

Different studies have administered OHIP with a reference period of one month. These include research on orofacial pain (Murray *et al.*, 1996; Segù *et al.*, 2005) and on patients

with oral rehabilitation needs (Awad *et al.*, 2000). The choice of this period, however, was not justified in either study.

To the best of my knowledge, there is no study in the literature that attempts to establish the optimal reference periods to assess TMD treatment outcomes.

2.5.3. Temporal dynamics

Some chronic pain conditions are characterised by evident variation in intensity over time, e.g. fibromyalgia (Gendreau *et al.*, 2003; Harris *et al.*, 2005). This can occur within the same day or from day to day. It is important, therefore, to take these dynamics into consideration when assessing treatment outcomes. Most studies assess treatment outcomes at different follow-up occasions with fixed or unfixed intervals (Farrar, 2000; Gendreau *et al.*, 2003). If this assessment looks only at the variable of interest at the time of the assessment, then it clearly misses what is happening between the review points. This is similar to taking snapshots of a moving object and expecting to understand how it moves. Some methods overcome this drawback by asking the patients to assess the outcome of interest over a period of time (the reference period). Yet, they fall into another problem, and that is relying on the patient's memory which often yields a bias in the assessment (Redelmeier and Kahneman, 1996).

Examining the temporal dynamics of a condition will not only enhance the assessment of treatment outcomes, but will improve understanding of the condition itself and its progress. This point is strongly emphasised by a review study of diaries (Bolger *et al.*, 2003).

Diaries are the most powerful tools to assess changes in the outcome of interest as they are happening (Bolger *et al.*, 2003). While this section aims only to highlight the importance of examining temporal dynamics of chronic disorders, assessment methods for these dynamics along with their pros and cons are covered elsewhere (section 2.4).

2.6. Assessment of oral health-related quality of life (OHrQOL)

OHrQOL measures concern the impact of orofacial conditions on quality of life. In a way, they can be considered as site-specific self-report measures of health.

The development of OHrQOL was inspired by the general concept of health-related QOL. Historically, OHrQOL was given little attention in comparison to generic HrQOL (Cohen and Jago, 1976). It was in 1984 that research was carried out on the societal impact of OHrQOL, more specifically the impact of dental diseases on loss of work (Reisine, 1984). Subsequent research has demonstrated that orofacial conditions have as serious an impact on QOL as other disorders (Cushing *et al.*, 1986; Reisine *et al.*, 1989).

The earliest attempts to measure OHrQOL used generic HrQOL instruments, for example, the Sickness Impact Profile (SIP) (Reisine, 1985). This approach was criticised by Locker who argued that while pain impact can be detected, the impact of other conditions such as tooth loss cannot (Locker, 1988). The first oral-specific measure of health related quality of life was the Sociodental Scale (Cushing *et al.*, 1986).

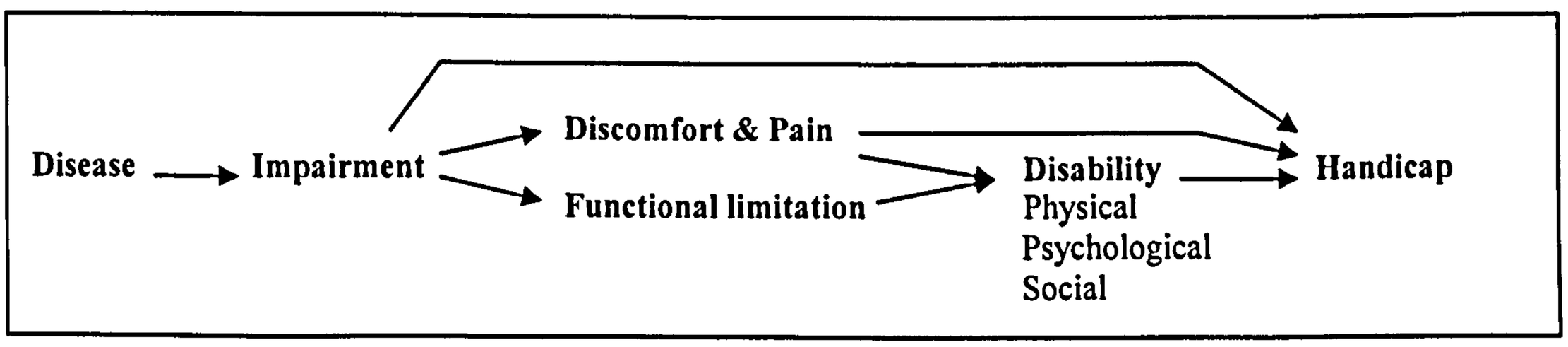
It is argued that measuring OHrQOL by generic HrQOL tools has a number of advantages over the use of OHrQOL specific measures. For instance, the use of generic instruments allows for the comparison between oral-related and other disorders. In addition, the psychometric characteristics of these measures are already established. The counter argument is that generic instruments often fail to capture subtle differences in oral health status (McGrath *et al.*, 2003), and that their discriminant validity and responsiveness to changes with treatment is poor (Allen *et al.*, 1999; Allen *et al.*, 2001). Furthermore, OHrQOL instruments are condition-specific, and therefore of less burden on the patient. This would consequently shorten completion time and improve the response rate.

2.6.1. Oral specific health related quality of life instruments

Similar to generic HrQOL, OHrQOL measures have been developed to measure the impact of illness both on the societal level, including days lost to sickness and treatment costs (Reisine, 1984) and on personal levels. Only the latter is considered here.

Different approaches and frameworks have been used to construct OHrQOL instruments. The conceptual framework based on WHO's classification of impairment, disability and handicap (World Health Organization, 1980) is a reasonable starting position (Sato, 1997; Allen, 2003). In this concept, any damage in the orofacial structures (e.g. tooth lost or TMJ arthritis) is considered as impairment. This can consequently lead to disability (e.g. difficulties in performing daily activities such as eating and speaking) and/or to handicap in more severe conditions (e.g. absence from work and financial consequences). Based on this classification, Locker built a frequently cited conceptual model for measuring oral health (Locker, 1988) (Figure 2.4).

Figure 2.4. Locker's conceptual model for measuring oral health



Locker's model forms the foundation for several OHrQOL instruments including: the Oral Health Impact Profile (OHIP) (Slade and Spencer, 1994), the Dental Impact on Daily Living (DIDL) (Leao and Sheiham, 1996), and the Oral Impacts on Daily Performances (OIDP) (Adulyanon *et al.*, 1996).

2.6.2. Historical uses of OHrQOL instruments

OHrQOL tools have been mostly used in descriptive population studies (Locker and Slade, 1993; Nuttall *et al.*, 2001; McGrath and Bedi, 2003a; Steele *et al.*, 2004; Fernandes *et al.*, 2006). Some workers established national norms for some of these tools (Kelly, 2000; McGrath and Bedi, 2002; John *et al.*, 2004b), while others have investigated the cross-cultural differences of OHrQOL (Slade *et al.*, 1996; Allison *et al.*, 1999).

OHrQOL instruments have also been used as treatment outcome measures (Rogers *et al.*, 1998) in clinical trials (Awad *et al.*, 2000; Allen and McMillan, 2003; Allen *et al.*, 2006) or as predictors for treatment needs (Leao and Sheiham, 1996; Locker and Jokovic, 1996). A few attempts have also been made to use these instruments as cost-utility indices (McGrath and Bedi, 2003b; Heydecke *et al.*, 2005; Brennan and Spencer, 2006). There is still some development work required for the use of such measures for outcome assessment and economic purposes.

2.6.3. Types of OHrQOL instruments

OHrQOL, with its different domains and dimensions, can be assessed using either single-item global rating question or multiple-item questionnaires. While the former is shorter, easier to administer and therefore more suitable for large scale cross-sectional research, the latter, having more items, is more accurate in capturing details and more appropriate for longitudinal studies. Nevertheless, multiple-item questionnaires have largely been used in surveys involving large number of participants (Slade *et al.*, 1996; Allen and McMillan, 2003; Steele *et al.*, 2004).

A number of instruments from both approaches (single and multiple-item) have been devised and used in the past. Single-item global ratings of oral health have been used as indicators of health status (Locker and Jokovic, 1996; Jones *et al.*, 2003). However, these do not attempt to measure QOL, for example, *"How would you describe the health of your teeth and gums?"*, which is asked with a five-point Likert type scale (from *Good* to *Poor*) (Jones *et al.*, 2003).

Multiple-item OHrQOL instruments have also been developed, with considerable heterogeneity in their focus, length and format (Table 2.10 page 93). For example, current OHrQOL questionnaires range in length between three and 56 items reflecting the divergence in opinions about definition of QOL and its components. Of these instruments, four allow patients to report the positive impact of oral health on QOL (Slade, 1997a; McGrath and Bedi, 2001; Slade, 2002). These are the Geriatric (or General) Oral Health Assessment Index (GOHAI), the Dental Impact Profile, the Oral Health Quality of Life – U.K. and the Oral Health QOL Inventory.

Condition-specific instruments

While OHrQOL instruments are considered as a site-specific type of the HrQOL instruments, even more condition-specific instruments have been developed as shortened versions of the general OHrQOL instruments [e.g. OHIP-edentulous (Allen and Locker, 2002)]. However, oral condition-specific measures have not all been derived by shortening generic OHrQOL tools. Some have been developed as original and independent OHrQOL instruments. Examples of these are the Xerostomia-Related Quality of Life Scale (XeQoLS) (Henson *et al.*, 2001), the Manchester Orofacial Pain Disability Scale (Aggarwal *et al.*, 2005), and the Liverpool Oral Rehabilitation Questionnaire (Pace-Balzan *et al.*, 2006).

Table 2.10. "Generic" oral health related quality of life instruments*

Instrument	Reference	Dimensions measured	Number of items	Response format
Sociodental Scale	(Cushing *et al.*, 1986)	Chewing, talking, smiling, laughing, pain, appearance	14	Yes/no
Geriatric (or General) Oral Health Assessment Index (GOHAI)	(Atchison and Dolan, 1990)	Chewing, eating, social contacts, appearance, pain, worry, self-consciousness	12	6 categories
RAND Dental Health Index	(Dolan *et al.*, 1991)	Pain, worry, conversation	3	4 categories
Dental Impact Profile	(Strauss and Hunt, 1993)	Appearance, eating, speech, confidence, happiness, social life, relationships	25	3 categories
Oral Health Impact Profile (OHIP)	(Slade and Spencer, 1994)	Function, pain, psychological discomfort, physical disability, psychological disability, social disability, handicap	49	5 categories
Subjective Oral Health Status Indicators (SOSHI)	(Locker and Miller, 1994)	Chewing, speaking, eating, communication, social relations	42	Various depending on the question
Dental Impact on Daily Living (DIDL)	(Leao and Sheiham, 1996)	Comfort, appearance, pain, daily activities, eating	36	Various
Oral Health-Related Quality of Life (OHQOL)	(Kressin *et al.*, 1996)	Daily activities, social activities, conversation	3	6 categories
Oral Impacts on Daily Performances (OIDP)	(Adulyanon *et al.*, 1996)	Performance in eating, speaking, oral hygiene, sleeping, appearance, emotion	9	Various
Oral Health Quality of Life – UK [OHQoL-UK(W)]	(McGrath and Bedi, 2001)	Physical, social and psychological dimensions	16	9 categories
Oral Health Quality of Life Inventory	(Slade, 2002)	Oral health, nutrition, self-rated oral health, overall QOL	56	4 categories
Child oral-health-related quality of life (CPQ_{11-14})	(Jokovic *et al.*, 2002)	Oral symptoms, functional limitations, emotional wellbeing and social wellbeing	36	Various

*** From (Jokovic *et al.*, 2002; Slade, 2002; Allen, 2003). Instruments are sorted in chronological order.**

2.6.4. Choosing an OHrQOL measure

Choice of the most appropriate OHrQOL measure relies on different considerations including the instrument's theoretical framework, purpose, ability to measure positive as well as negative impact of health on QOL, derivation method, psychometric characteristics, practicality and cross-cultural compatibility.

To date, no substantive work has been published to compare the performance of the different available OHrQOL instruments (Allen, 2003). Some studies, however, have partially explored this issue. For example, Strassburger *et al.* found that OHIP is increasingly used in OHrQOL studies, and that it is methodically more sufficient than other OHrQOL instruments (Strassburger *et al.*, 2004).

The instrument of choice should have a sound theoretical basis that is in accordance with the domains of quality of life and health-related quality of life. OHIP is structured around Locker's hierarchal model of health and its related quality of life (Locker, 1988), which is based in turn on the WHO's classification of illness.

Existing OHrQOL instruments have been developed for various purposes, as discussed in section 2.4.3.3.The psychometric characteristics of an instrument are dependent on the purpose it was designed for. Therefore, it is essential to choose the instrument for the right purpose and with suitable properties. OHIP was originally developed to measure the impact of oral disorders on the individuals in epidemiological research. Nevertheless, it has been employed successfully in different types of research including its use as a measure of treatment outcomes for prosthetics treatments (Awad *et al.*, 2000; Allen and McMillan, 2003), surgical interventions (Ross Beirne, 2004), orthodontics (de Oliveira and Sheiham, 2004), and tooth whitening (McGrath *et al.*, 2005). OHIP has also been used in research on oral conditions such as Behçet's disease (Mumcu *et al.*, 2006), periodontal diseases (Ng and

Leung, 2006) and temporomandibular disorders (Murray *et al.*, 1996; John *et al.*, 2002; Larsson *et al.*, 2004; Segù *et al.*, 2005; Wolfart *et al.*, 2005).

On the other hand, unlike other measures such as the Oral Health Quality of Life – UK (McGrath and Bedi, 2001), OHIP only allows for negative rating of health (Locker and Gibson, 2006). In other words, it measures deviation from health and the burden of illness as opposite to measuring well-being. Also, it measures only frequency of the impact on quality of life and not its severity.

OHIP's major advantage over some other instruments is that it is based on statements derived from in-depth interviews with a representative sample of dental patients (Slade and Spencer, 1994). This gives it an excellent content validity. In addition, OHIP's psychometric properties have been extensively explored and proven in different conditions and populations (Slade, 1997b; Slade, 2002; John *et al.*, 2003; Larsson *et al.*, 2004). Furthermore, OHIP appeared to be the only OHrQOL instrument tested or used in the TMD population.

Practicalities

In the search for a practical and inexpensive measure of OHrQOL, it is particularly important that the instrument can be successfully self-administered, thus eliminating the costs and resources often needed for the interviewer-administered questionnaires. Robinson *et al.* (2001) compared between OIDP (Slade and Spencer, 1994) and a short version of OHIP-49: OHIP-14 (Slade, 1997b) (see section 2.6.6 below). Both instruments were administered in two formats, interviews and self-completed questionnaires. The authors found that OHIP-14 preformed better than OIDP in the self-completed format. In part, this may be due to the fact that OIDP was originally designed and tested as an interviewer-administered instrument. The completion rate of the self-completed OIDP was particularly low in ethnic minorities, who apparently found it difficult to complete the questionnaire (Robinson *et al.*, 2001).

This conclusion is supported by findings from another study that compared OHIP-14 and OIDP in a sample of patients with xerostomia (Baker *et al.*, 2006). However, OHIP suffers a limitation in comparison to OIDP, and that is it does not measure severity of the problem but frequency.

When compared to the Geriatric (or General) Oral Health Assessment Index (GOHAI) (Atchison and Dolan, 1990), OHIP-14 was found to be as good a discriminative tool, but weaker in detecting within-subject change (Locker *et al.*, 2001). This may be explained by the way OHIP-14 was derived from the original OHIP-49. The derivation method means that OHIP-49 lost a large number of items, and with them it lost part of its psychometric properties, as discussed in section 2.6.6.

In addition, with the increasing international collaboration of medical research, OHrQOL instrument that are cross-culturally compatible have advantage on others which are not (Bullinger *et al.*, 1993; John *et al.*, 2007). Similarly, instruments which can be used for different conditions, allow for cross-condition comparisons, something that is very helpful for health bodies and policy makers.

Partly as a result of its early development and perhaps of its strong theoretical framework, OHIP has gained international popularity and has been translated to different languages including Chinese, German, Hebrew, Hungarian, Italian, Japanese, Malay, Spanish and Swedish. The international popularity of OHIP gives it even more strength as it allows for comparisons to be made not only across different nations, but also across the different disease populations for which it has been used. For example, Larsson *et al.* (2004) used the Swedish OHIP to compare four conditions: TMD, primary Sjögren's Syndrome, oral mucosal pain and skeletal malocclusion.

2.6.5. The Oral Health Impact Profile (OHIP)

OHIP comprises 49 statements structured in seven domains which are matched more or less to Locker's model (Locker, 1988). These are functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability and handicap (Appendix 7). It does not measure, however, the perceived satisfaction with oral health, changes in oral health, prognosis or self-reported diagnosis (Slade, 1997a).

An example of the basic style of OHIP items is: *Have you found it uncomfortable to eat any foods because of problems with your jaws, teeth, mouth or dentures?* Each item is scored on a five point scale related to frequency of impact in the following order: *Never, Hardly ever, Occasionally, Often, and Very often.* The reference period in the original OHIP is one year, i.e. patients assess their quality of life retrospectively over one year.

Weights are available for OHIP items based on Thurstone's method of paired comparisons (Allen and Locker, 1997; Slade, 1997a). These aim to attribute to the items values that are related to their importance. However, it is argued that these weights unnecessarily make OHIP scoring more complicated and unfeasible in clinical settings (Allen and Locker, 1997; Allen *et al.*, 2001). In most current research employing OHIP weights are not used.

Scoring OHIP

Calculating OHIP scores is achieved in two ways. The first is to compute summed item scores "Sum OHIP". So for a particular individual an overall score is computed by summing item scores. For a group, a mean score can be obtained for each independent item across all members of the group. Since the scale is ordinal and not continuous, overall scores are appropriate only for comparison purposes, and not as summary scores. In other words, a score of 20 represents more impact on quality of life than that of 10, but the absolute value itself is meaningless as it could refer to different degrees of the impact frequencies.

The second method is to count items answered at a threshold level, e.g. fairly often or very often (FOVO). Individuals' overall scores are calculated based on the number of items that reach this threshold. For a group, the overall score for a particular item is the proportion of respondents reporting this item above the given threshold. Statistically, this is simpler and more straightforward to handle than the previous method. However, in many populations the distribution of this summary variable may be skewed, with many respondents reporting no impact at the defined threshold, restricting the use of parametric statistics (Slade, 2002).

2.6.6. Short-form versions of OHIP

Different distilled versions of OHIP have been produced in order to facilitate its use in clinical or epidemiological settings, or to obtain disease-specific measures (Allen, 2003). Historically, OHIP was first reduced by Slade to a 14-item version, the so called OHIP-14 (Slade, 1997b). This was followed by another attempt to derive a 14-item version using a different shortening method (Locker and Allen, 2002). These latter authors then produced a condition-specific version of OHIP, OHIP-EDENT for the edentulous population (Allen and Locker, 2002). This instrument is also known as OHIP-20.

The simplicity and usefulness of the short versions of OHIP prompted authors around the world to produce their own short form including the Chinese OHIP-14S, the Malaysian, the Japanese, the Hebrew 14-item OHIP. Recently the Germans introduced three new daughters of OHIP: the 5-item, the 14-item and the 21-item (John *et al.*, 2006b).

Different methods have been used to derive the short forms including regression analysis, factor analysis and item-impact analysis (as discussed in section 2.4.3.5). The choice of shortening method depends on three factors:

1. Whether the original instrument is the gold standard for the condition it measures
2. Whether the desired short-form is intended to replace the long form in the same context, or to be a population (disease) specific version
3. Whether the intended short-form is desired to act as a discriminative or evaluative tool.

Slade used internal reliability analysis, factor analysis and regression analysis to reduce OHIP and found that the regression analysis yielded the optimum set of items (OHIP-14) (Slade, 1997b). Locker and Allen adopted a different approach to obtain a 14-item version by using a modified form of the item-impact analysis (Locker and Allen, 2002). In the original item-impact analysis (Juniper *et al.*, 1997) each item is given a score, which is obtained by multiplying its prevalence in the population of interest by a measure of its importance. The modified form used mean score of the item as a proxy for its importance. Items were then ranked by their item-impact and the top two items from each OHIP dimension were selected.

Locker and Allen compared their results with Slade's and found that items derived with the regression analysis have a smaller size effect and a greater floor effect than those obtained by the item-impact method (Allen and Locker, 2002; Locker and Allen, 2002). They concluded that Slade's regression analysis yielded a good discriminatory tool but that was weaker in describing the OHrQOL of populations and detecting change with treatment.

Furthermore, Locker and Allen (2002) reported that both the regression and item-impact methods produced indices with good psychometric characteristics, including discriminant and convergent validity, internal consistency, and correlation with the full OHIP. While the regression analysis did better in some of these tests, the authors argue that any subset of 14 items of OHIP will have reasonable psychometric properties no matter how they are selected (Locker and Allen, 2002). Therefore, the approach of choice to shorten OHIP depends on the ultimate use of the index.

2.6.7. OHIP as a measure of TMD-related quality of life

Section 2.2.6 of this review reported the impact of TMD on patients' quality of life, and the different patient-based instruments that have been used to measure this impact. Among these was OHIP, which was able to capture the impact of TMD and orofacial pain on sufferers' lives and to differentiate between TMD sufferers and the general population. (Murray *et al.*, 1996; John *et al.*, 2002; Larsson *et al.*, 2004; Segù *et al.*, 2005). For example, TMD patients reporting frequent or constant function-related problems such as chewing food were four times as many compared to those in the community, and nine times more reported psychological problems such as depression (Murray *et al.*, 1996). OHIP was also sensitive to the differences between different diagnostic groups of orofacial pain (Murray *et al.*, 1996), and also sensitive to changes with treatment (John *et al.*, 2002).

Among these studies, only Murray *et al*'s study used the English language OHIP. Murray *et al* pioneered the use of OHIP in the TMD population providing a very useful insight on the extent to which quality of life was compromised in these patients and how this was associated with different pain profiles. The significant correlation between OHIP and other pain-related measures run in parallel support the validity and reliability of OHIP as a measure of TMD-related QOL.

That study, however, suffered from some design limitations, which were recognised by the authors. For example, the patients were selected based on their clinical notes, and were mailed the questionnaires prior to their initial appointments. Clinical examination to confirm the diagnosis was only available for half of the participants. Moreover, the sample consisted of patients with TMD and other neurological pain disorders. According to the authors, these groups differ in their quality of life profile (different OHIP domain scores were noted). Furthermore, only 30 items of OHIP were used in this study; 21 items were eliminated and the choice of these questions was subjective, and two items were added to OHIP: *"Taking longer to complete a meal"* and *"Avoiding eating with others"* based on previous research

undertaken by the authors. Finally, the one-year reference period in the original OHIP was changed to one month, with no justification for this time frame.

In the German version of OHIP "OHIP-G" (John *et al.*, 2002) the authors tested the ability of OHIP to differentiate between TMD sufferers and four other groups including a group of healthy subjects. They also tested OHIP's ability to capture change with treatment. The extensive analysis concluded that OHIP-G was both effective as a discriminative and evaluative tool for TMD. The authors also reported a strong correlation between health-related quality of life assessed by OHIP and RDC/Axis II measures (John *et al.*, 2007).

Despite the relatively small TMD patient sample (n = 30) that took part in validating the Swedish version (Larsson *et al.*, 2004), the study reported interesting findings. These authors added the word *"jaws"* to the source of patients suffering in the body of OHIP questions. The discriminative psychometric properties for OHIP were examined in five groups including subjects coming for routine dental check-ups, patients with TMD, primary Sjögren's Syndrome, oral mucosal pain and skeletal malocclusion. Additionally, the authors confirmed the stability of OHIP in the TMD subgroup.

The Italian version of OHIP was also validated in a sample of TMD patients (Segù *et al.*, 2005). The study was to a large extent a reproduction of Murray *et al.*'s study (1996). The authors used the same 30 OHIP items, same one month reference period, and same validating pain measures. This work confirmed the findings of the above-mentioned and OHIP was found to have good discriminative psychometric properties for TMD.

All studies that have used OHIP in TMD populations reported that non TMD-related items dilute the overall score. They agreed that exclusion of these items would not only minimize respondent burden, but more importantly would improve the performance of OHIP as a measure of TMD-related quality of life.

2.7. Summary of the literature review

Emphasis has been increasingly given to base the understanding of TMD aetiology, their management and outcome assessment on the same biopsychosocial model used for other chronic pain conditions (Dworkin and Massoth, 1994; LeResche, 1997a).To this end, outcome measures are needed for TMD that encompass this concept and that are based on patients' perception of their illness and of their quality of life as opposed to clinician-based measures.

Currently, most TMD measures concern about the detectable physical signs and symptoms. An exception to that is the RDC system (Dworkin and LeResche, 1992), which allows assessment of both the physical and psychological dimensions. However, RDC is clinician-centred, does not assess the social dimension of TMD and has not been tested as an outcome measure.

Other patient-based non TMD-specific instruments are available, which measure the impact of chronic illness in its multidimensional model. However, these generic instruments are arguably not sensitive enough to capture subtle differences in health status (Hill *et al.*, 1996; Turk *et al.*, 2003). In this perspective, balance is required between condition-specific instruments that are sensitive to small changes with treatment and generic health instruments which permit comparisons across different conditions, a pre-requisite for decision-making in health care. To this end, OHrQOL measures provide a reasonable balance and can form the basis for the development of a TMD outcome measure.

Among available OHrQOL measures, the Oral Health Impact Profile (OHIP) (Slade and Spencer, 1994) is the most widely used both across various oral conditions and different languages. Moreover, OHIP has a strong theoretical framework, has been extensively validated and has shown ability to capture the impact of TMD on quality of life.

OHIP, however, has not been tested as an outcome measure for TMD. Its many non TMD-related items might potentially compromise its utility as an outcome measure.

Moreover, its one-year reference period, set originally for epidemiologic studies on oral diseases, is not suitable for outcome assessment of TMD.

A research is needed to identify OHIP items that are most relevant to TMD, which exhibit sufficient psychometric properties to form the core of a patient-based quality of life outcome measure for TMD. An investigation is also required to establish the optimal time to administer OHIP to assess TMD treatment outcome and to determine the best reference period to be used in this assessment. This background forms the rational for this PhD project.

Chapter 3.
Aims of project

The overall aim of this project was to start the development of a quality of life outcome measure for TMD. Three linked studies were conducted for this purpose. The aims of each of these studies are shown below:

Study One: Determination of suitable review and reference periods to assess TMD treatment outcomes: Analysis of TMD pain diaries

Pain dynamics of patients undergoing conservative treatment of TMD were explored to establish optimal review and reference periods for the assessment of TMD treatment outcome.

Study Two: Determination of suitable review and reference periods to assess TMD treatment outcomes: A qualitative approach – Interviews with TMD patients

This study aimed to triangulate with *Study One* and verify the previously identified review and reference periods. Qualitative methods were used to explore the lived experience of TMD throughout treatment, and to understand patients' perception of the process of improvement.

Study Three: Establishing a quality of life measure of TMD treatment outcomes

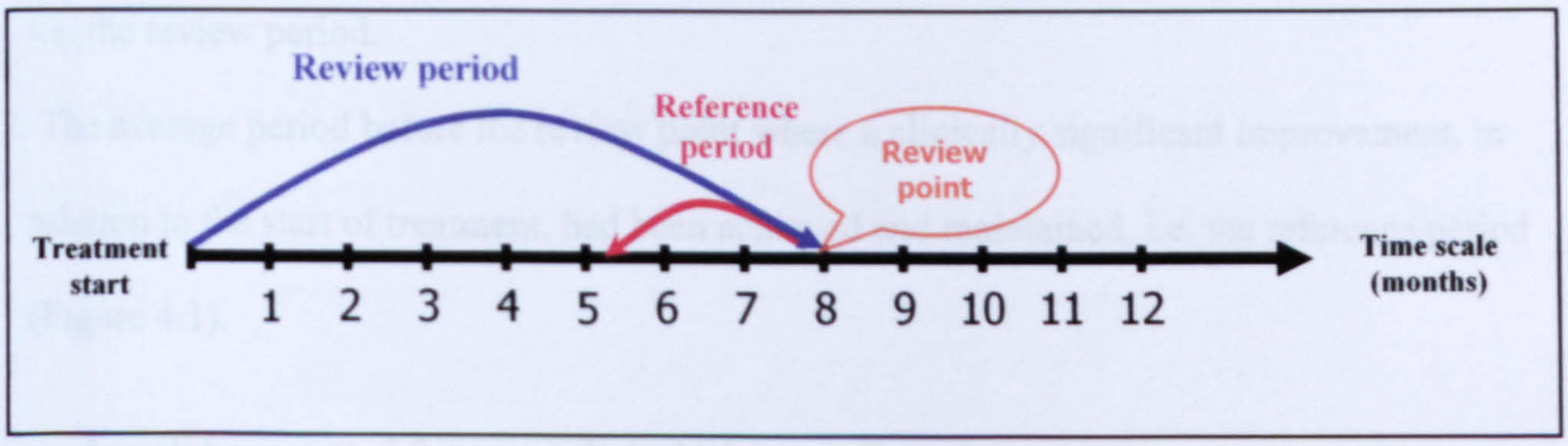
Study three aimed to begin the process of developing an outcome measure for TMD, using as a starting point the Oral Health Impact Profile (OHIP). The study aimed to derive a short form of OHIP that is potentially suitable as an outcome measure and to explore aspects of its psychometric properties.

Chapter 4.
Study One: Determination of suitable review and reference periods to assess TMD treatment outcomes:
Analysis of TMD pain diaries

4.1. Introduction

To conduct reliable and proper assessment of treatment outcomes, sufficient time must be allowed for the treatment to work. This is critical when different approaches are being compared. In this part of the thesis, *Study One*, I aimed to determine an appropriate time to review TMD treatment outcomes. I called the time at which a meaningful assessment of treatment outcomes is made the *"Review Point"*, and the time between the start of treatment and this point is therefore called the *"Review Period"* (Figure 4.1). In the literature, this period has been chosen arbitrarily and varies from one month to one year (Forssell *et al.*, 1999).

Figure 4.1. A time scale representing review and reference periods



The treatment outcome assessment to be carried out, in this project using the Oral Health Impact Profile (OHIP), a quality of life instrument, will ask patients to appraise their conditions retrospectively over a period of time, called the *"Reference Period"* (Figure 4.1).

The concept of reference period appears only rarely in the TMD literature. OHIP, was originally designed for patients with oral and dental problems in general and used a reference period of 12 months (Slade and Spencer, 1994; Slade, 1997b). However, John *et al.* (2002) found that a reference period of one month provided OHIP with the best internal consistency. Some studies that administered OHIP on orofacial pain patients used a one month reference period (Aggarwal *et al.*, 2005; Segù *et al.*, 2005). These latter authors, however, did not justify their choice of one month.

To the best of my knowledge there is no literature that determines when TMD treatment outcomes can be measured properly (i.e. the treatment has been given sufficient time to work), and I aimed from this study to determine suitable review and reference periods to assess TMD treatment outcomes. To achieve this aim I analyzed changes in pain intensity with treatment of TMD patients. The data consisted of pain diaries recorded by TMD patients during a trial of splint therapy (Wassell *et al.*, 2004). Changes in pain scores throughout the treatment were used to draw a general "pain recovery" trend for those who responded to treatment. This trend was used then to inform:

1. The average time taken for TMD treatment to reach and maintain an acceptable outcome, i.e. the review period.
2. The average period before the review point where a clinically significant improvement, in relation to the start of treatment, had been achieved and maintained, i.e. the reference period (Figure 4.1).

As will be expected from any clinical trial, not all participants respond to treatment and this was the case with patients in this study (as explained later). However, as aforementioned, the patients of interest were only those who showed improvement (the *Improvers*). Therefore, I had first to discriminate improvers from non-improvers; and to do so I had to design some discrimination rules and compare them against other traditional measures.

I will call the preparatory study which explored pain dynamics and discriminated between improvers and non-improvers: *Study 1-A*; and the following work which aimed to study the pattern of pain improvement in an average improver: *Study 1-B*. Since the materials (i.e. the patient group that provided the pain diary data) are the same for both studies, I will describe these first, and then I will give details for the methods and results for each study separately.

4.2. Aim

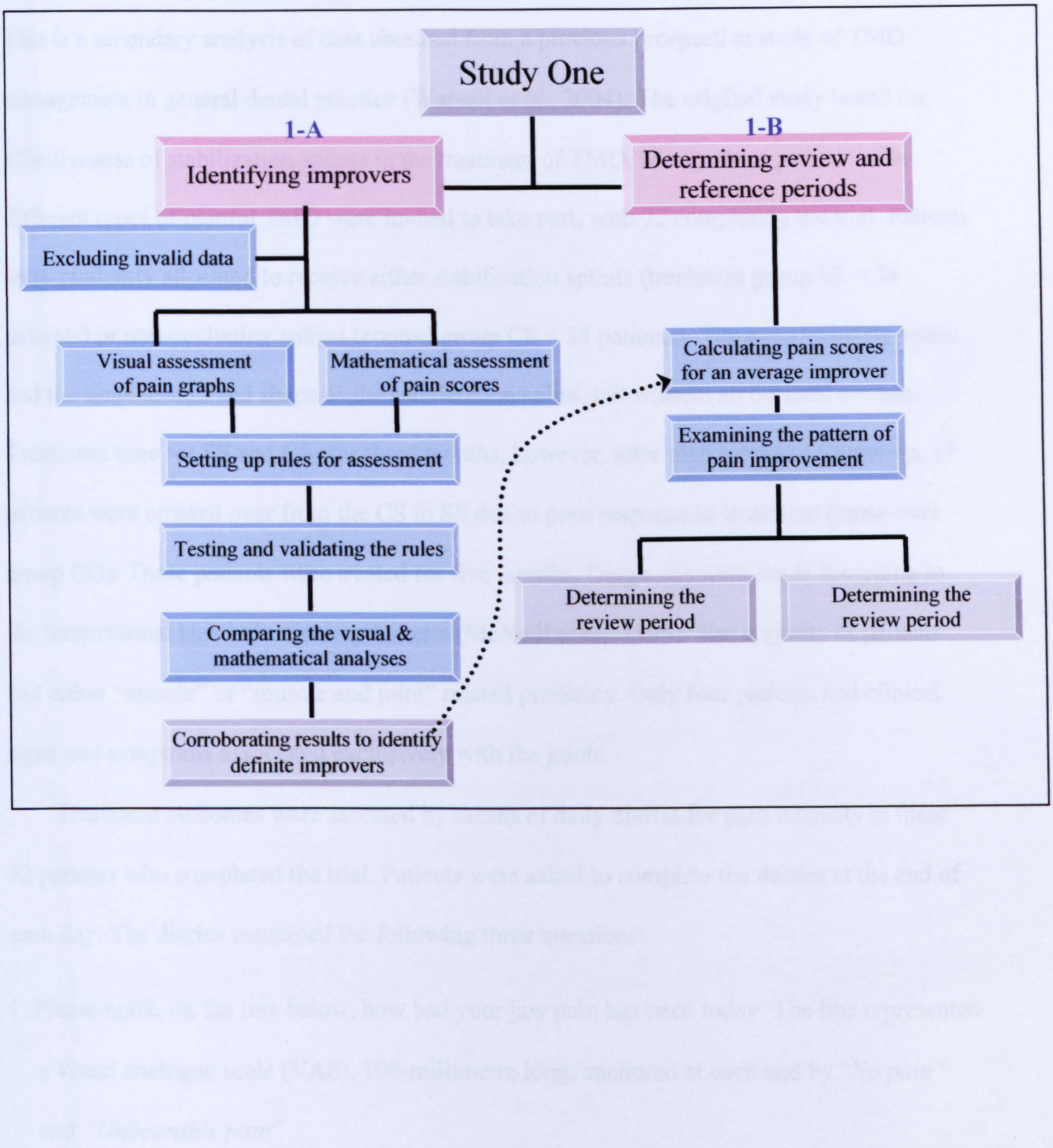
This study explored pain dynamics of patients undergoing conservative treatment of TMD, with the aim of establishing suitable review and reference periods to assess TMD treatment outcome.

4.3. Objectives

1. Study 1-A: to identify the improvers by:
    a. Assessing characteristics of pain change with time by visually examining pain/time plots
    b. Setting up rules to discriminate patients based on the above characteristics
    c. Testing the reliability of these rules
    d. Comparing these rules against other traditional mathematical measures of improvement
    e. Identifying the definite improvers to inform Study 1-B
2. Study 1-B: to examine the general pattern of pain improvement, in order to determine:
    a. A suitable review period
    b. A suitable reference period.

Outline of this study is presented in Figure 4.2.

Figure 4.2. Outline of *Study One*



4.4. Materials

This is a secondary analysis of data obtained from a previous prospective study of TMD management in general dental practice (Wassell *et al.*, 2004). The original study tested the effectiveness of stabilization splints in the treatment of TMD. Ninety three patients with different types of painful TMD were invited to take part, with 72 completing the trial. Patients were randomly allocated to receive either stabilization splints (treatment group SS = 34 patients) or non-occluding splints (control group CS = 38 patients). The non-occluding splint had the same design and shape as the stabilization splint, but without an occlusal surface. Treatment time for SS and CS was three months, however, after nine weeks of follow-up, 17 patients were crossed over from the CS to SS due to poor response to treatment (cross-over group CO). These patients were treated for five months. Diagnoses were made according to the International Headache Society Criteria (McNeill *et al.*, 1993). The majority of patients had either "muscle" or "muscle and joint" related problems. Only four patients had clinical signs and symptoms associated exclusively with the joints.

Treatment outcomes were assessed by means of daily diaries for pain intensity in these 72 patients who completed the trial. Patients were asked to complete the dairies at the end of each day. The diaries contained the following three questions:

1. Please mark, on the line below, how bad your jaw pain has been today. The line represented a visual analogue scale (VAS), 100-millimetre long, anchored at each end by "*No pain*" and "*Unbearable pain*"

2. How many hours of headache have you suffered today?

3. How many painkillers have you taken today? ☐

   Type of tablets? ----------

4.5. Study 1-A: Identifying improvers

Objectives of this study are summarized in page 109.

4.5.1. Methods

Pain VASs were measured using a millimetre ruler and scores to the nearest millimetre inserted into an Excel spreadsheet. Some of the diaries were not valid for two reasons:

1. Five patients had more than one diary for the same period of time. These unexplained double data were excluded.
2. Not all patients filled in the diaries regularly, and these were excluded if they had missing data for:
   a. More than 10% overall, or
   b. More than 25% missing data within any month (any 30 consecutive days)

This resulted in a further exclusion of 28 patients. Therefore, 39 patients met the inclusion criteria, representing 54% of patients completing treatment. These were 34 females and five males with mean age 35 years (ranging from 19 to 65 years). Sixteen of them were in the SS group, 13 controls (CS) and 10 were crossed-over from SS to CS. Detailed diagnoses using the International Headache Society Criteria (McNeill *et al.*, 1993) are shown in Table 4.1.

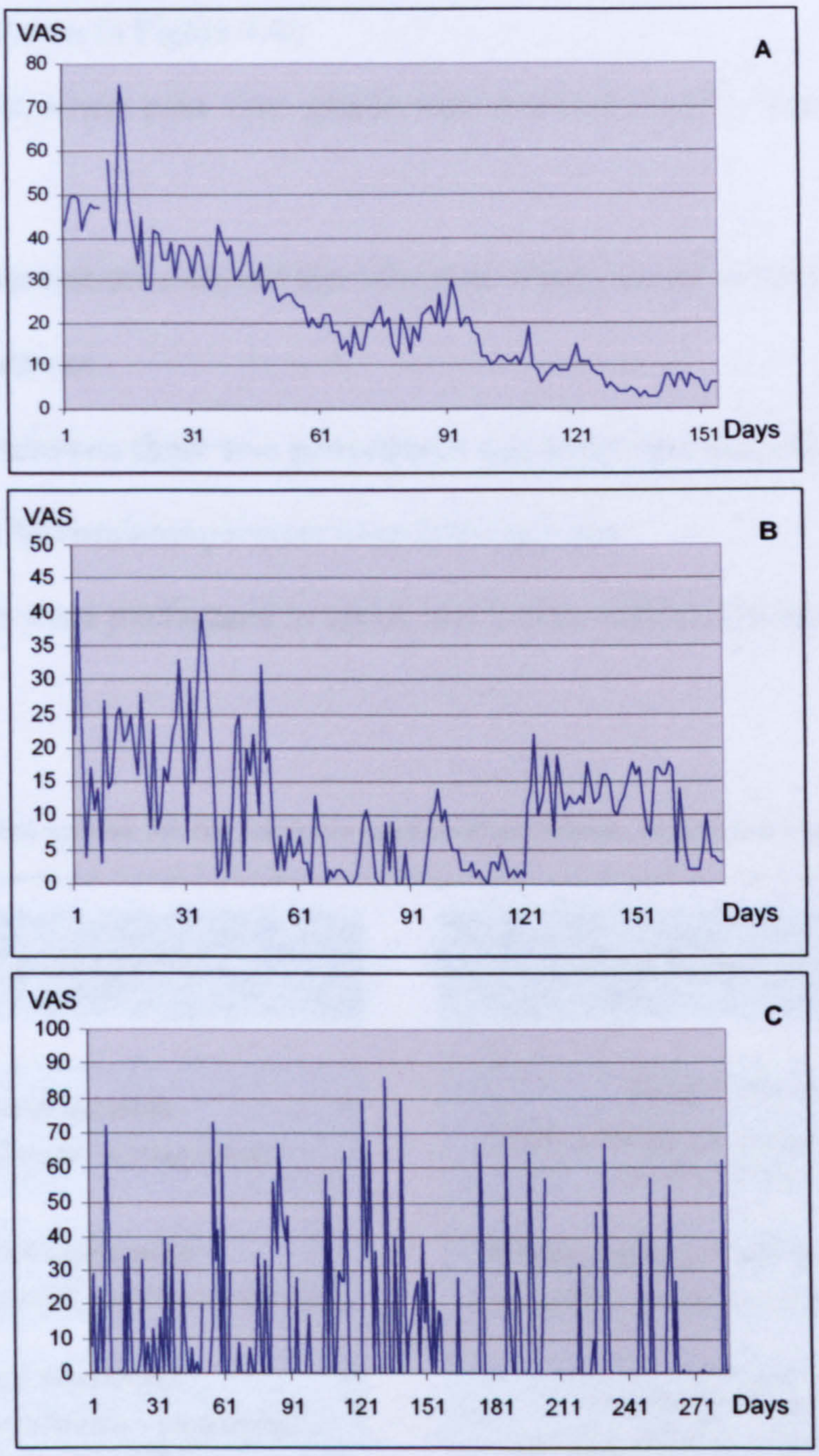
Table 4.1. Diagnoses of patients participating in the original trial in *Study One*

	Sub-diagnosis	Frequency	%
Muscle Disorders	None	1	2.6
	Reflex Splinting	14	35.9
	Myofascial Pain	24	61.5
	Total	39	100.0

	Sub-diagnosis	Frequency	%
Joint Disorders	None	23	59.0
	Disc Displacement with Reduction	15	38.5
	Capsulitis	1	2.6
	Total	39	100.0

Thirty nine individual pain/time graphs were drawn which revealed remarkable dissimilarity among patterns of pain change with treatment. Some graphs showed a clear trend in jaw pain reduction during treatment, others showed no improvement, but only pain fluctuation, whilst a third group was borderline between improvers and non-improvers. Examples of these graphs are shown in Figure 4.3.

Figure 4.3. Different types of pain/time change with treatment

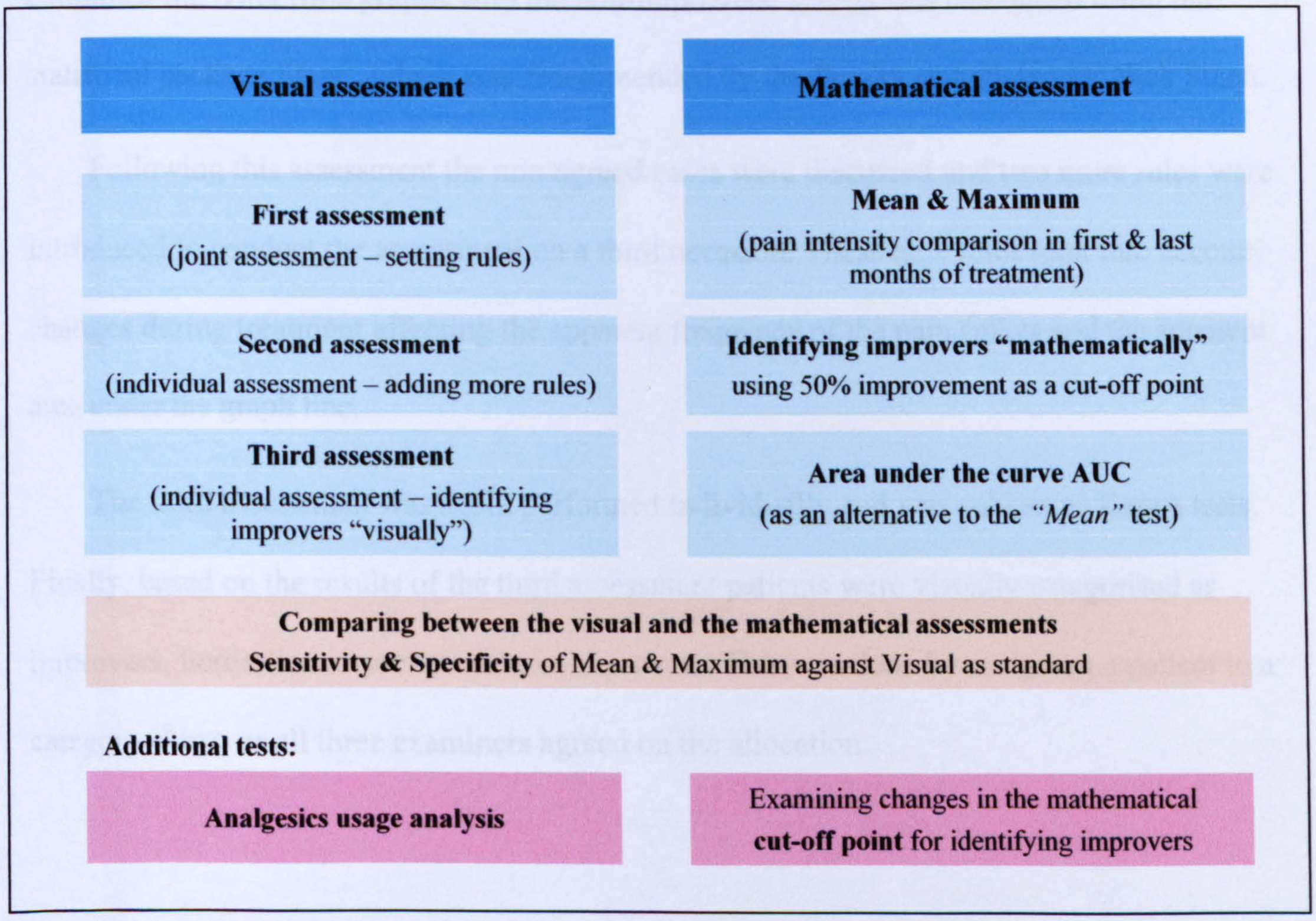


A. Improver
B. Borderline improver
C. Non-improver

To identify definite improvers, these pain-time graphs underwent the following analyses (detailed outline is shown in Figure 4.4):

1. **Visual assessment**, where pain–time graphs were assessed visually; assessment rules were set and tested.
2. **Mathematical assessment** analysed the reduction of pain scores in relation to the first and last months of treatment.
3. **The agreement between these two procedures** was tested and definitive responding patients identified by combining results from both analyses.
4. **Further analyses were performed** to check and further validate the aforementioned approach.

Figure 4.4. Detailed outline for the methods used to discriminate improvers from non-improvers



4.5.1.1. Conducting and validating the visual assessment

Graphs plotted for the selected 39 patients were examined by three consultants experienced in the management of TMD, two of whom were specialists in Restorative Dentistry and the third in Oral & Maxillofacial Surgery. To improve and validate the assessment criteria, this was conducted on three separate occasions with two month-intervals to reduce memory effects. In the first assessment the assessors jointly examined the pain/time graphs and considered two factors: 1- general trend of the graph, 2- heights of the pain spikes (Figure 4.3).

To determine the inter-examiner reliability of the visual assessment, on the second occasion the consultants assessed the graphs individually. Their agreement was measured by two sets of Kappa tests. The first considered all three patient categories, whilst the other combined the borderline graphs with the non-improvers. Kappa was calculated using the statistical package Stata[1], which was recommended by the faculty statistician Dr Nick Steen.

Following this assessment the non-agreed cases were discussed and two more rules were introduced to conduct the assessment on a third occasion. These new rules took into account changes during treatment affecting the apparent frequency of the pain spikes and the apparent area under the graph line.

The third assessment was again performed individually and was subject to Kappa tests. Finally, based on the results of the third assessment patients were visually categorised as improvers, borderline improvers or non-improvers. This was done by assigning a patient to a category if two or all three examiners agreed on the allocation.

[1] Stata 7, StataCorp LP, 4905 Lakeway Drive, College Station, Texas 77845, USA

4.5.1.2. The mathematical assessment

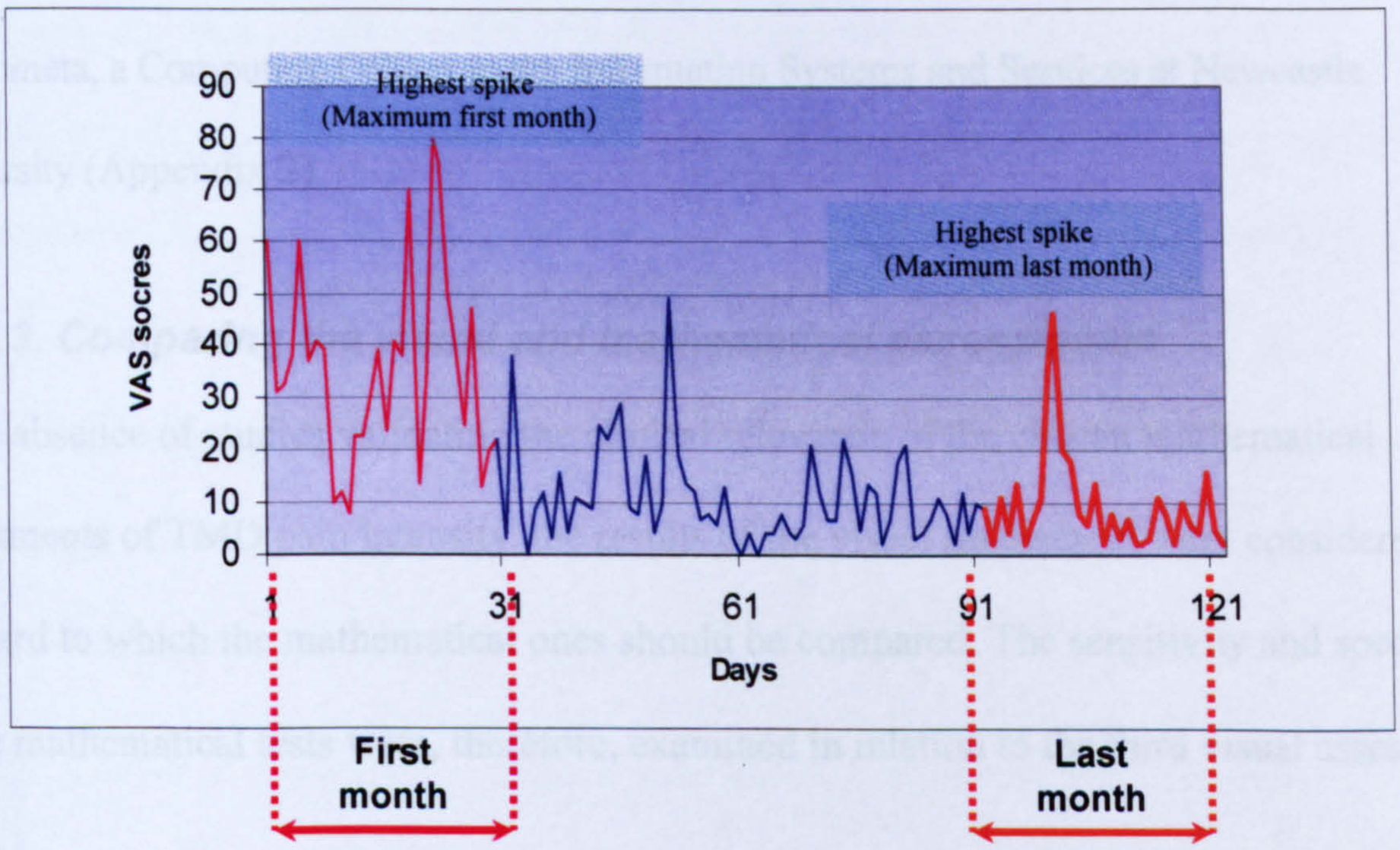
The second approach to discriminate improvers from non-improvers analysed changes of pain scores during treatment mathematically. Two measures were considered: the *Mean* of pain scores over a month and *Maximum* pain score in a month (Figure 4.5).

To be considered an improver, a patient must have achieved both:

1. A 50% reduction of the first month mean pain score at the last month of treatment
2. A 50% reduction of the first month maximum pain score at the last month of treatment

These rules were derived in collaboration with the study statistician Dr Nick Steen. The agreement between the two rules was measured by means of a Kappa test. Then, the results obtained from both mathematical approaches were combined and compared with those resulting from the visual assessment.

Figure 4.5. A plot showing the time periods involved in the mathematical assessment of mean pain, AUC and height of maximum spike (an example of one of the improvers assessed visually)



Area under the curve as an alternative to the mean pain scores

The relationship between the *"Mean"* pain scores and the commonly used measure "*area under the curve, AUC"* was examined.

To calculate *AUC* it was first necessary to replace the few items of missing data. The statistical software Statistical Package for the Social Sciences (SPSS)[1] suggests four methods to replace missing values, however, two of them, regression and EM estimation (expectation-maximization), "depend on the assumption that the pattern of missing data is related to the observed data only. This assumption allows estimates to be adjusted using available information". These two methods were adopted to replace the missing values.

For each patient *AUC* was measured for first and last month, and a 50% reduction in *AUC* (first month compared to last month) was used as a cut-off point to discriminate improvers from non-improvers. The Correlation Coefficient (*r*) was calculated for the comparison of the results obtained by *AUC* and those obtained by the *"Mean"* test.

Having replaced the missing values, *AUC* was calculated using a formula advised by Dr ST Kometa, a Computing Officer in the Information Systems and Services at Newcastle University (Appendix 2).

4.5.1.3. Comparing the visual and mathematical assessments

In the absence of studies validating the clinical relevance of the chosen mathematical assessments of TMD pain intensity, the results of the visual assessment were considered the standard to which the mathematical ones should be compared. The sensitivity and specificity of the mathematical tests were, therefore, examined in relation to the third visual assessment.

[1] SPSS help section, SPSS 12.0.1 FOR Windows. Release (11 Nov 2003). SPSS Inc. Headquarters, 233 S. Wacker Drive, 11th floor , Chicago, Illinois 60606.

4.5.1.4. Identifying the definite improvers

Since the aim of *Study One* was to analyse data of definite improvers, only those cases with both visual and mathematical evidence of improvement were selected.

4.5.1.5. Optimum threshold for discriminating improvers from non-improvers

The 50% threshold of pain reduction to discriminate improvers from non-improvers in the mathematical test was chosen as it has been adopted arbitrarily by a number of pain studies. To determine the effect of changing the 50% threshold I also used 40% and 60% reductions and examined the sensitivity and specificity of each mathematical discrimination against the visual one. The threshold providing the highest aggregate sensitivity and specificity was identified as optimum.

4.5.1.6. Analgesics usage analysis

The approach described above to identify improvers is based on professionals' perception of pain data. To help validate this approach I used a measure of patients' response to pain, namely consumption of self-prescribed analgesics used as an additional treatment to support splint therapy. It was hypothesised that only the improvers group would show a significant reduction in the usage of analgesia when comparing the first and last months of treatment.

Patients were asked to record the type of analgesic and number of tablets taken. However, to avoid problems with making comparisons of different self-prescribed analgesics I analysed the number of days each patient used analgesics regardless of the analgesic type or amount of tablet consumed.

The number of days when analgesics were used by improvers and non-improvers during the first and last months of treatment were compared using the Kruskal-Wallis Test. Individual comparisons within groups with time and between groups at each time period were performed with Mann-Whitney U Test. The level of significance was $p<0.05$.

4.5.2. Results

4.5.2.1. Conducting and validating the visual assessment

The first visual examination of the graphs resulted in the identification of three groups. While Appendix 3 presents all graphs, examples of the three groups are shown in Figure 4.3.

1. **Improvers:** Patterns where there was a clear trend in pain reduction (20 patients)
2. **Non-improvers:** Patterns with no specific trend to reduction or increase, but with major fluctuations giving a saw tooth appearance of varying frequency (13 patients) or patterns where pain remained constant (one patient)
3. **Borderline improvers:** Patterns showing a possible trend to pain reduction but accompanied by several spikes of pain (five patients)

The second assessment, carried out individually, resulted in total agreement in 24 out of the 39 cases (61%), and partial disagreement (two out of three examiners agreeing) for the remaining 15 cases. Examiners ratings of each graph are presented in Appendix 3.

Two Kappa tests were performed: *Kappa 2-A* measured the agreement between the three assessors for the three categories; *Kappa 2-B* measured the agreement between the three assessors over definite improvers. In other words, non-improvers and borderline improvers were combined in one group as being non-improvers. The results were *Kappa 2-A* =0.6 and *Kappa 2-B* = 0.76. This time 18 improvers, seven borderline improvers and 14 non-improvers were identified according to agreement between either two or three examiners.

The third assessment observed a general improvement in agreement with total agreement in 29 cases and partial agreement for 10 cases only (Appendix 3). Similar Kappa tests were conducted and the results were *Kappa 3-A* =0.70 (3rd assessment / three categories) and *Kappa 3-B* = 0.79 (3rd assessment / two categories). On this occasion 20 improvers, nine borderline improvers and ten non-improvers were identified. Table 4.2 compares the results obtained from all three visual assessments, and shows a trend to an increase in borderline cases with repeated assessment.

Table 4.2. Comparison between results from the three visual assessments

Category	First assessment	Second assessment	Third assessment
Improvers	20	18	20
Borderline improvers	5	7	9
Non-improvers	14	14	10

The detailed Kappa analysis showed consistently good agreement for identification of improvers and non-improvers, but identification of borderline improvers was much less reliable (Table 4.3).

Table 4.3. Agreement between examiners (Kappa test) for the second and third assessments

Category	Second assessment Kappa 2-A	Third assessment Kappa 3-A
Improvers	0.76	0.79
Borderline improvers	0.26	0.44
Non-improvers	0.69	0.79

In relation to the type of treatment provided, Table 4.4 shows similar proportions of patients in each TMD treatment group classified visually as improvers, borderline and non-improvers.

Table 4.4. Cross-tabulation of TMD treatment group with visually assessed treatment outcome

		Visual Assessment			
		Improvers	**Borderline**	**Non improvers**	**Total**
Trial type	**Control**	7	2	4	13
	Stabilization	9	5	2	16
	Crossed-over	4	2	4	10
	Total	20	9	10	39

4.5.2.2. The mathematical assessment

Unlike the visual assessment the calculations did not allow for borderline improvers to be assigned, so patients were classified as improvers or non-improvers only. There was only moderate agreement between the *"Mean"* and the *"Maximum"* tests (Kappa=0.58). Combining the results from these tests 21 improvers and 18 non-improvers were identified.

The *"Area under the curve, AUC"* test identified exactly the same improvers and non-improvers as did the *"Mean"* test. The correlation coefficient, ***r***, for the correlation between *AUC* and *"Mean"* was 0.99. Therefore, *AUC* was not included in further analyses.

4.5.2.3. Comparing the visual and mathematical assessments

When examined individually against the visual assessment, the *Mean* test and the *Maximum* test had sensitivities of 0.9 and 0.84 respectively but the specificity for both was only 0.68. When the results of the *Mean* and *Maximum* tests were combined the sensitivity remained at 0.9 but the specificity improved to 0.84.

4.5.2.4. Identifying the definite improvers

The corroboration between results from both the visual and mathematical tests resulted in the identification of 18 improvers as shown in Table 4.5.

All the improvers were female, aged between 19 and 59 years with a mean age of 34.5 years. Nine of them received the stabilization splint, while six were controls and three were crossed-over.

Table 4.5. Corroboration of the visual and mathematical assessments to derive definite improvers*

Visual	Mathematical	Combined	Definite Improvers
1	1	1	1
4	4	4	4
6	6	6	6
10	10	10	11
11	11	11	13
13	13	13	14
14	14	14	15
15	15	15	16
16	16	16	17
17	17	17	18
18	18	18	21
21	21	21	24
23	23	23	25
24	24	24	26
25	25	25	30
26	26	26	32
30	30	30	33
32	32	32	34
33	33	33	
34	34	34	
2	2	2	
8	8	8	
9	9	9	
12	12	12	
19	19	19	
20	20	20	
29	29	29	
35	35	35	
38	38	38	
3	3	3	
5	5	5	
7	7	7	
22	22	22	
27	27	27	
28	28	28	
31	31	31	
36	36	36	
37	37	37	
39	39	39	

Legend

- Improvers
- Borderline
- Non-improvers

*** Numbers represent patients' trial numbers**

4.5.2.5. Optimum threshold for discriminating improvers from non-improvers

Table 4.6 shows the results for the sensitivity and specificity tests comparing the visual and the mathematical assessments carried out with different thresholds. The optimum percentage of pain reduction used as a threshold to discriminate between improvers and non-improvers was 50%, above and below which the aggregate sensitivity and specificity of the mathematical assessment dropped. This was true for the *Mean*, *Maximum* and combined tests.

Table 4.6. Sensitivity & specificity of the mathematical assessment against the third visual assessment with different cut-off points

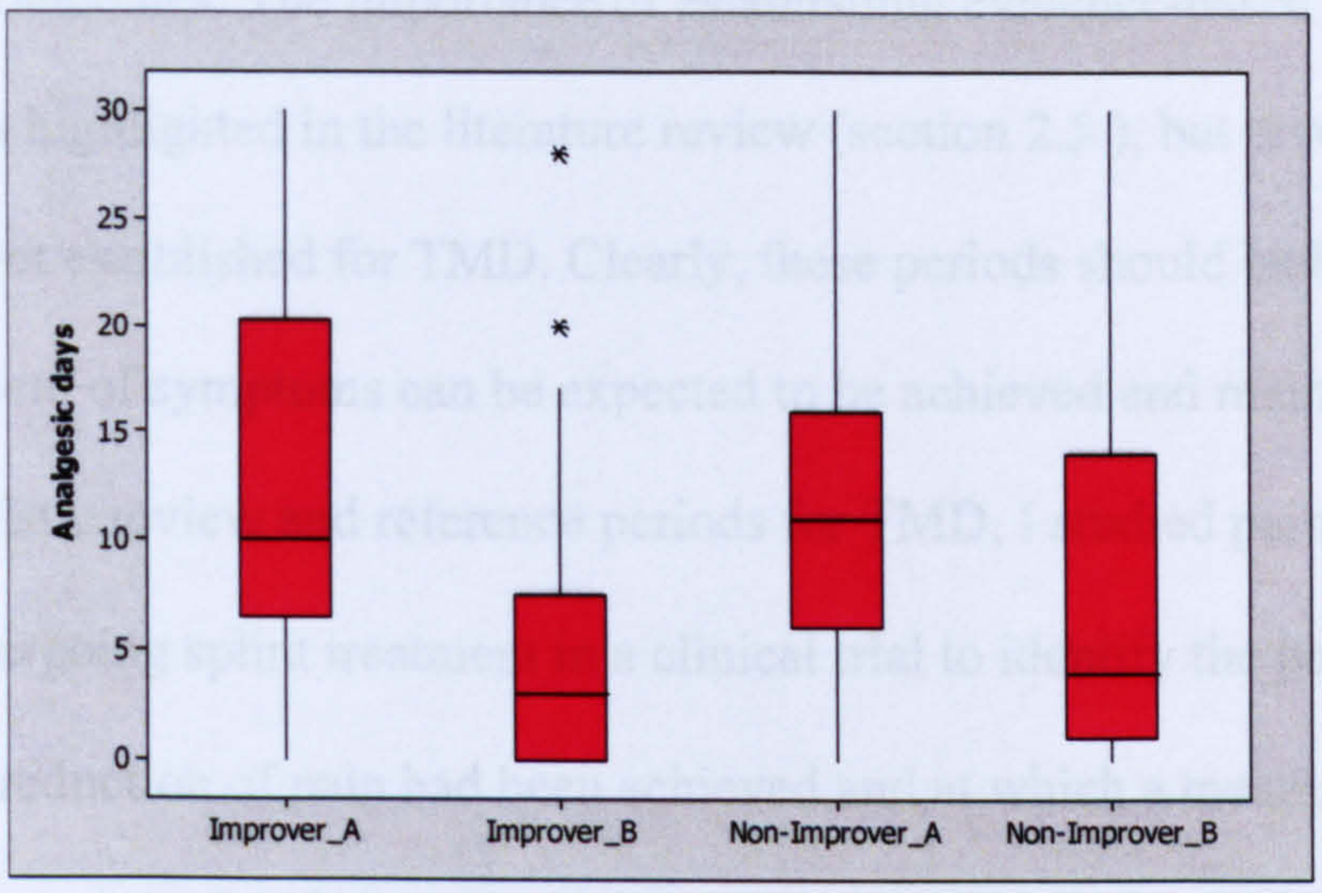
		Non improver if, compared to first month, remaining pain >		
		40%	**50%**	**60%**
Mean	**Sensitivity**	0.85	0.95	0.95
	Specificity	0.74	0.68	0.58
	Aggregate	1.59	1.63	1.53
Maximum	**Sensitivity**	0.80	0.90	0.95
	Specificity	0.74	0.68	0.63
	Aggregate	1.54	1.58	1.58
Combined	**Sensitivity**	0.75	0.90	0.95
	Specificity	0.84	0.84	0.74
	Aggregate	1.59	1.74	1.69

4.5.2.6. Analgesics usage analysis

Both improvers and non-improvers had very similar numbers of days when analgesics were taken during the first month of treatment. Only four improvers and one non-improver used no analgesia during the course of the trial. Kruskal-Wallis test demonstrated a significant difference ($p=0.02$) between the four groups (improvers and non-improvers, in the first and last month) as shown in Figure 4.6. Patients classified visually as improvers ($n=20$) showed a significant reduction in median number of analgesic days when comparing the first and last

months (10 c.f. 3, p=0.01, Mann-Whitney U test). There was a trend to a reduction in analgesic days with the non-improvers (n=19) but this was not significant (p=0.09).

Figure 4.6. Number of days analgesics used during treatment



A: First month
B: Last month

4.5.3. Discussion

The original aim of this study was to establish suitable review and reference periods for TMD treatment outcome assessment. These periods were necessary to examine the performance of the intended quality of life questionnaire (the core of this project – *Study 3*) as a tool for measuring treatment outcomes. The importance of establishing evidence-based review and reference periods was highlighted in the literature review (section 2.5), but revealed that this period had not been yet established for TMD. Clearly, these periods should be based on when significant improvement of symptoms can be expected to be achieved and maintained.

To establish realistic review and reference periods for TMD, I studied pain/time changes of TMD patients undergoing splint treatment in a clinical trial to identify the point at which a clinically significant reduction of pain had been achieved and at which a meaningful assessment of improvement might be made by patients. I was only interested in those patients who showed improvement in their symptoms, the so called *Improvers*. If I had not separated out the improvers, the pain/time responses of the non-improvers would have masked how much improvement could be expected and how long it took.

It was first essential to find a valid method to identify these improvers *(Study 1-A)*. At a later stage, data of an average improver was examined in detail *(Study 1-B)* to establish the review and reference periods.

To discriminate improvers from non-improvers, pain/time graphs from 39 patients were visually analysed. Criteria were set to identify the graphical characteristics of improvers, and were tested for validity and enhanced. The raw data were then re-examined mathematically in order to express the visual criteria in arithmetical terms that are often used in reporting treatment outcomes. Having identified the improvers, more analyses were carried out to validate the results. This involved triangulating the results with the proportion of improvers in the treatment groups and with analgesic use.

In this study, I analysed daily changes in pain intensity and considered the full course of treatment. I favoured this approach over relying on discrete measures weeks or months apart because in chronic conditions these are unreliable. In chronic conditions such as TMD, pain is often fluctuating from day to day (Gendreau *et al.*, 2003). Discrete measurements, therefore, often mislead as they may capture high or low points on the graph that are not necessarily representative. This point was highlighted in published recommendations to avoid bias in VAS analysis (Gendreau *et al.*, 2003). Indeed, discrete measures simply represent snapshots of a continuous fluctuating course of changes. To compensate for this drawback, some studies tend to assess the pain over a retrospective period of time. But these measures are in turn prone to bias due to distorted recall memory effects (Eich *et al.*, 1985; Bolger *et al.*, 2003). This is why pain diaries are the favoured tool to capture changes in self-reported pain in a way that is closer to real time than other methods (Stone and Shiffman, 1994; Bolger *et al.*, 2003).

In discussing the results two caveats must be emphasised in relation to TMD. Firstly, TMD consists of a variety of sub-diagnoses the most common of which are myofascial pain and disc displacement with reduction. These conditions are well represented in the present sample (Table 4.1). For sub-diagnoses occurring less frequently (e.g. disc displacement without reduction and TMJ arthritis) the pain/time characteristics may well be different. Secondly, in determining overall treatment response for TMD, as for other musculoskeletal problems, outcomes other than pain intensity need to be considered. These include clinical findings (e.g. muscle and joint tenderness, clicking, and limitation of opening), quality of life measurement, analgesic consumption and the need for rescue medication (Farrar *et al.*, 2003).

4.5.3.1. The visual assessment

To achieve robust results, strict criteria were implemented to exclude cases with unacceptable missing data that might affect the assessment. Following that, 39 graphs were plotted, which revealed major variations not only between patients, but also for individual patients on a day

to day basis. These findings agree with those reported in studies looking at fibromyalgia (Gendreau *et al.*, 2003; Couper, 2005).

Experienced clinicians are in a unique position to make informed judgements about what pattern of pain dynamics constitutes a meaningful improvement, particularly in the context of a condition that they treat regularly. The ability of clinicians to agree independently is a pre-requisite though. Criteria for visual assessment were refined to optimise agreement between observers as determined by Kappa. The simple general trend (slope and diminishing area under the curve) of the graph was neither adequate to fully describe a pattern of response, nor sufficient to obtain good agreement between the assessors. Additional criteria included apparent frequency and height of pain spikes. Reduction in pain spikes has been recommended recently as part of the core outcome measures in chronic pain (Aaron *et al.*, 2005). The final criteria yielded the best agreement between the three assessors (Kappa = 0.79).

The assessors were relatively consistent in identifying the same improvers and non-improvers. However, the borderline improvers were more challenging. These cases showed a comparably poor level of agreement between the assessors (Table 4.3), which noticeably affected their overall agreement (Kappa).

4.5.3.2. The mathematical assessment

In this assessment I measured mean pain scores, maximum pain score and area under the curve *(AUC)* for first and last month of treatment. This helped express the visual criteria in mathematical terms that are often used in reporting treatment outcomes. The first and last month of treatment were chosen to simplify analysis; some patients had taken three months or less to show clinically significant improvement, whilst others had taken up to five months of treatment. In relation to the overall length of treatment a one month sampling period provided a

satisfactory measure of pain/time reduction at the end of treatment. It is worth mentioning here that the monthly period suits TMD, but might not be appropriate for other disorders.

The *Mean* value represented an aggregate value of pain changes over the one month period. This method has been used often in pain research and is better than asking patients to retrospectively assess their pain over a period of time to avoid memory effects (Bolger *et al.*, 2003). Changes in the *Mean* were expected to represent the graph slope.

AUC, which has also been used before (Gendreau *et al.*, 2003), was expected to represent the corresponding visual criteria, i.e. "apparent area under the curve". *AUC* is believed to be more meaningful than the *Mean* pain score, and possibly is so when there is a considerable amount of missing data as it interpolates between existing data. In this study, however, when strict rules were applied to exclude subjects with missing data, the *AUC* and mean pain values identified exactly the same improvers and non-improvers. Therefore, *AUC* was not included in further analyses.

Finally, maximum pain score was chosen to represent the level of highest spike. Neither the *"Mean"* nor the *"Maximum"* test alone was sufficient to describe and report the process of pain reduction assessed visually by expert clinicians.

4.5.3.3. Comparing the visual and mathematical assessments

By combining results from the *Mean* and *Maximum* mathematical tests I achieved optimum agreement with the visual assessment in terms of specificity and sensitivity. These results confirm that statistical analyses should be used carefully and only within a clinically relevant frame to be meaningful.

As regards the best cut-off point, I found that a 50% reduction in mean and maximum pain was the best threshold to discriminate improvers from non-improvers. By establishing this cut-off point I have established the "Clinically Important Difference" CID for TMD disorders (see section 2.5.1).

4.5.3.4. Further validation – the analgesic consumption test

In order to further validate the judgement from a patient's perspective it is possible to triangulate the definition of success against an objective proxy for the patient's pain experience. In this case analgesic intake was used as a patient-based measure. Although a fairly blunt instrument, a significant reduction in analgesic usage was found in the improver group ($p=0.01$). The trend to a reduction in analgesic days in the non-improver group suggests that some of this group were also showing signs of improvement, even though the pain levels had reduced by less than 50%, but as this included some cases originally classified as borderline it is perhaps not surprising. In order to minimise the risk of type two errors, a strict definition of improvement is appropriate and it was this that underpinned the grouping of borderline cases with non-improvers.

It was possible to use analgesic days as a separate measure of pain experience in this group as 87% of patients were regularly using analgesics at the start of treatment. In other patient populations where subjects may avoid analgesia this type of comparison would not be possible. Nevertheless, the consumption of 'rescue medication' has been used as an outcome measure in clinical trials (Farrar *et al.*, 2000; Bolger *et al.*, 2003).

4.5.3.5. Further validation – comparing with trial results

Another factor in support of the validity of the visual definition was the similarity in the proportion of improvers in each of the splint groups. During the clinical trial there was no significant difference between the splint groups for any of the clinical outcome measures (Wassell *et al.*, 2004).

4.5.3.6. Strengths and limitations

The main strength of this study is that it challenged the traditional methods to identify improvers. I have closely observed the full course of pain changes throughout treatment and have used various outcomes (including a form of rescue medicine) to propose a new method to discriminate improvers from non-improvers.

Identifying improvers was critical not only to establish the review period, but also to comply with the recent trend in the medical literature. Analysis of pain measurements in clinical trials has moved away from consideration of mean pain changes of treatment and comparison groups, which can be misleading, towards comparing the proportions of patients in such groups who have shown a clinically significant improvement (Farrar *et al.*, 2000). The argument is that "group mean differences could reflect large changes in a few patients, small changes in many patients, or any combination of these outcomes" (Farrar *et al.*, 2000; Gendreau *et al.*, 2003).

Furthermore, this study enables the estimation of CID (the clinically important difference) for TMD using a different approach to that used in the medical literature. This could potentially support other research on CID for chronic pain conditions.

Although reported for the management of TMD, this approach is straightforward and could be applied to any condition where there is a risk of daily variation in the recovery period. This study does not advocate simply using clinical judgement as an outcome in clinical trials, but proposes applying clinicians' judgements to reliable patient data to make sure that any statistical definition of success has some clinical meaning, rather than being arbitrary.

On the other hand, this study was limited firstly by the lack of true pre-treatment data. Ideally, in any pain study the base-line evaluation of pain should take place before any intervention. Unless waiting lists are used constructively for this purpose however, there are ethical difficulties in withholding treatment to facilitate measurement. In this study pain

diaries were issued after the initial consultation which occurred two weeks prior to starting splint treatment. Whilst it is tempting to label this period 'pre-intervention', in reality the initial interaction between patient and clinician may have influenced pain perception so these were not truly pre-treatment measurements.

Secondly, the study was limited by the return of adequately completed pain diaries. In this study 54% of the 72 patients who finished treatment had adequately completed diaries. This provided sufficient data for the purpose of this methodological study; however, it would have been a disappointing result if the original trial (Wassell *et al.*, 2004) had relied entirely on the diaries as the sole outcome measure. Much better compliance with pain diaries has been reported by other workers (LeResche *et al.*, 2003; Aaron *et al.*, 2004). LeResche paid her patients to participate and also telephoned them as a reminder to complete the diaries. Aaron used electronic diaries to prompt patients to complete the diaries and to monitor their compliance. Clearly, if pain diaries are to be adopted as a principal outcome for chronic pain trials considerable efforts will be needed to ensure adequate compliance. Indeed, the lack of reporting of patient compliance and the impact of missing data have been raised as significant issues affecting the quality of many clinical trials of pain management.

Thirdly, the original study used paper type diaries. In contrast to electronic diaries that have become available recently, paper diaries suffer not only from lower compliance but from retrospective completion. This exposes the diaries to the risk of recall bias, the very drawback they have been used to avoid (Bolger *et al.*, 2003). One way to reduce this risk is to ask patients to submit their diaries at each follow-up visit. This study and others have used this method (LeResche *et al.*, 2003), nevertheless, the electronic diaries are still favoured in that they log time for completion, and can allow a pre-defined margin of delay (Bolger *et al.*, 2003; Gendreau *et al.*, 2003).

Fourthly, while the original study used different pain outcome measures, all were uni-dimensional. In other words, they all concerned pain intensity but not other pain measures such as verbal description, unpleasantness, coping, etc.

Finally, data about the rescue medicine (i.e. analgesic usage in addition to splint treatment) were available and were used as a secondary outcome to confirm previous findings. However, I was not able to use this as a gold standard. This is because assessment of pain was made at the end of the day. Patients' scores for "how bad your pain was today" can be interpreted as the overall level of pain before or having taking the medicine. Other workers have asked their patients to assess pain just before they take the rescue medicine (Farrar *et al.*, 2000). The challenge of using rescue medicine data in chronic disorders was indeed highlighted by one of the review studies (Farrar, 2000).

4.5.4. Summary

I examined the full course of TMD pain changes throughout conservative treatment. Different patterns of changes of pain scores were described. Some showed improvement while others did not or were not conclusive.

I established and validated methods to identify improvers using pain graphs and traditional statistics. Fifty percent reduction in pain was found to be a suitable cut-off point [or clinically important difference (CID)] to discriminate between improvers and non-improvers.

Eighteen improvers were identified when combining the visual and mathematical assessments. Data of the patients will be examined in the following study to determine the best time for outcome assessment.

4.6. Study 1-B: Examining the general pattern of improvement

The purpose of this study was to determine the review and reference periods for TMD treatment outcome. Having established patients showing improvement in *Study 1-A*, I examined in *Study 1-B* the pattern of pain changes for an average improver to identify a point where clinicians can make a meaningful outcome assessment (review point), and the period over which patients might make a retrospective assessment of their condition (reference period).

The review point was based on the time when improvement reached a plateau. The same concept has been used in TMD before (Pehling *et al.*, 2002), but without a clear definition of "improvement". The reference period is the time extended from the review point backward, to which patients will refer when making a self-report assessment. This period should be long enough for the patients to make meaningful assessment of their conditions. On the other hand, it should not extend too long so it reaches the stage when a clinically important improvement "CID" had not been achieved.

4.6.1. Methods

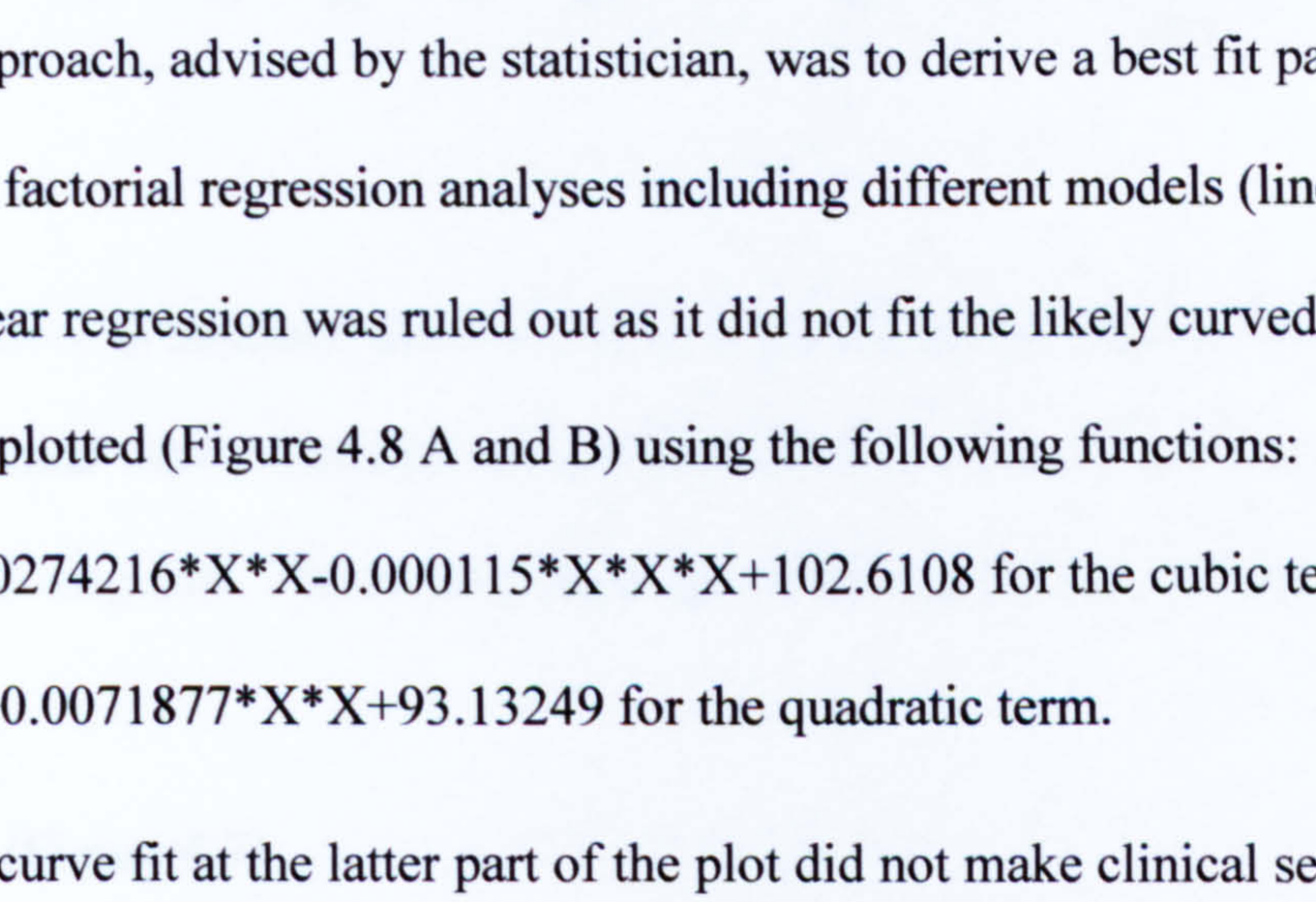
Two methods were used to examine the combined improvers' data. In the first I plotted the average daily pain (the mean daily pain scores across all improvers). The second method involved a regression analysis of the improvers' data, in an attempt to obtain a general trend of improvement.

It was first necessary to convert the VAS scores for each patient into a percentage of the pain at the treatment start (the first score). This allowed percentage of improvement (change of pain) for each individual to be examined rather than the actual pain score. This conversion was approved by the study statistician.

The period over which the mean pain improvement was calculated started from the first visit two weeks before the splints were fitted (Figure 4.7). Since treatment periods were not

exactly the same for all participants, means were calculated up to the point when 80% of the patients had completed treatment which was 126 days, i.e. four months. Data after this period were excluded to avoid corrupting results with the data from a few exceedingly slow responders.

Figure 4.7. Pattern of daily pain changes for an average improver



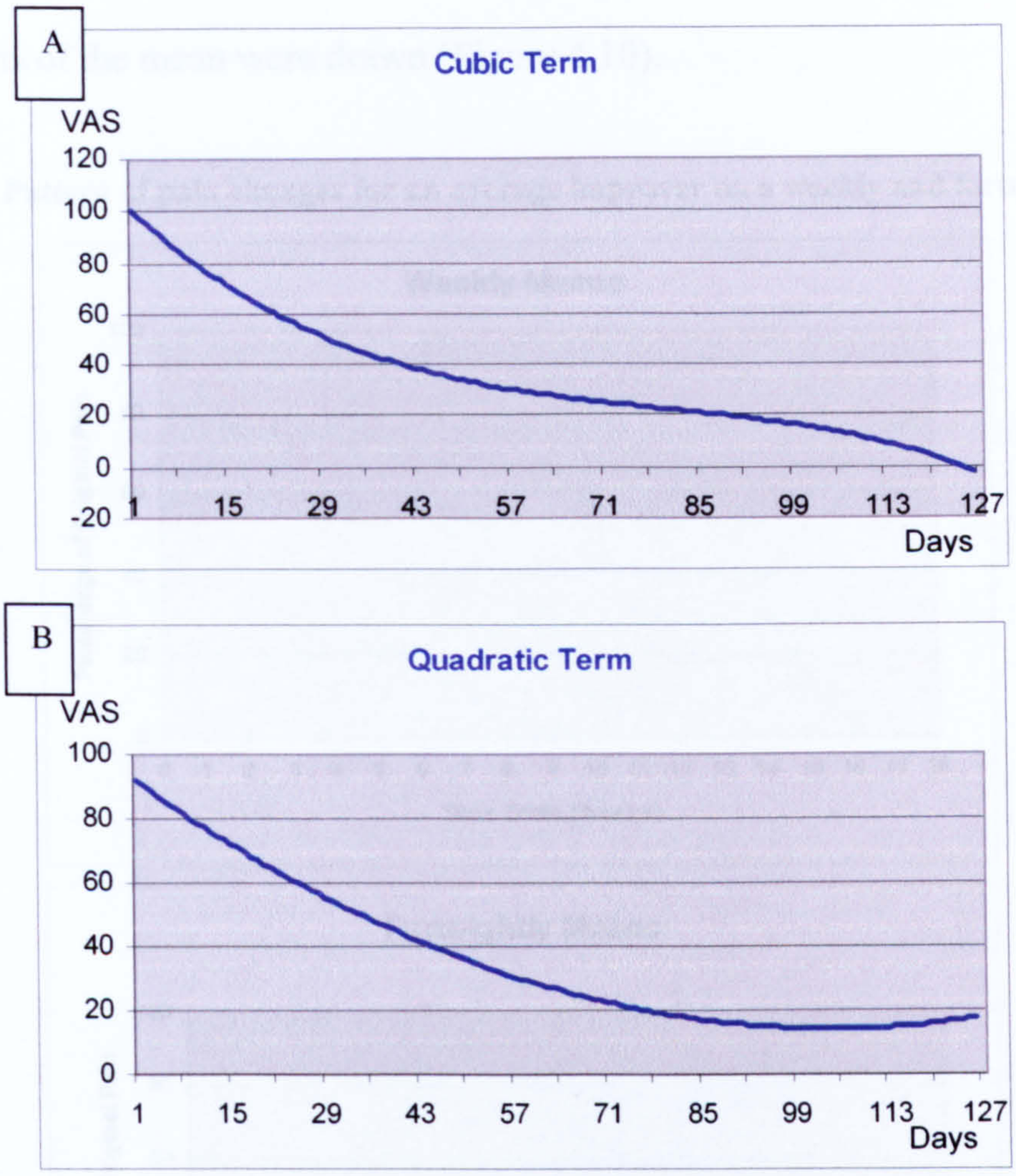
The second approach, advised by the statistician, was to derive a best fit pain/time response involving factorial regression analyses including different models (linear, cubic and quadratic). The linear regression was ruled out as it did not fit the likely curved plot. The other models were plotted (Figure 4.8 A and B) using the following functions:

Y = -2.4583*X+0.0274216*X*X-0.000115*X*X*X+102.6108 for the cubic term

Y = -1.505424*X+0.0071877*X*X+93.13249 for the quadratic term.

However, the curve fit at the latter part of the plot did not make clinical sense showing negative or increased values of pain. The simpler approach of using means calculated for selected time periods was therefore adopted.

Figure 4.8. Regression analyses of pain scores for an average improver



I decided therefore to continue the analysis with the simple plot representing aggregated improvers' data. Means were calculated using Microsoft Excel taking into account cells in the spreadsheet with missing data. The mean percentage of change for the 18 improving patients was plotted:

1. On a daily basis (Figure 4.7)

2. On a weekly basis (Figure 4.9)

3. On a fortnightly basis (Figure 4.9)

The daily graph was considered too spiky for the determination of the review and reference periods. On the other hand, too much detail was lost on the 2-weekly graph. As a result, I opted to complete the study with the weekly graph.

To check the graph's accuracy, standard errors of the weekly means were calculated and confidence limits of the mean were drawn (Figure 4.10).

Figure 4.9. Pattern of pain changes for an average improver on a weekly and fortnightly basis

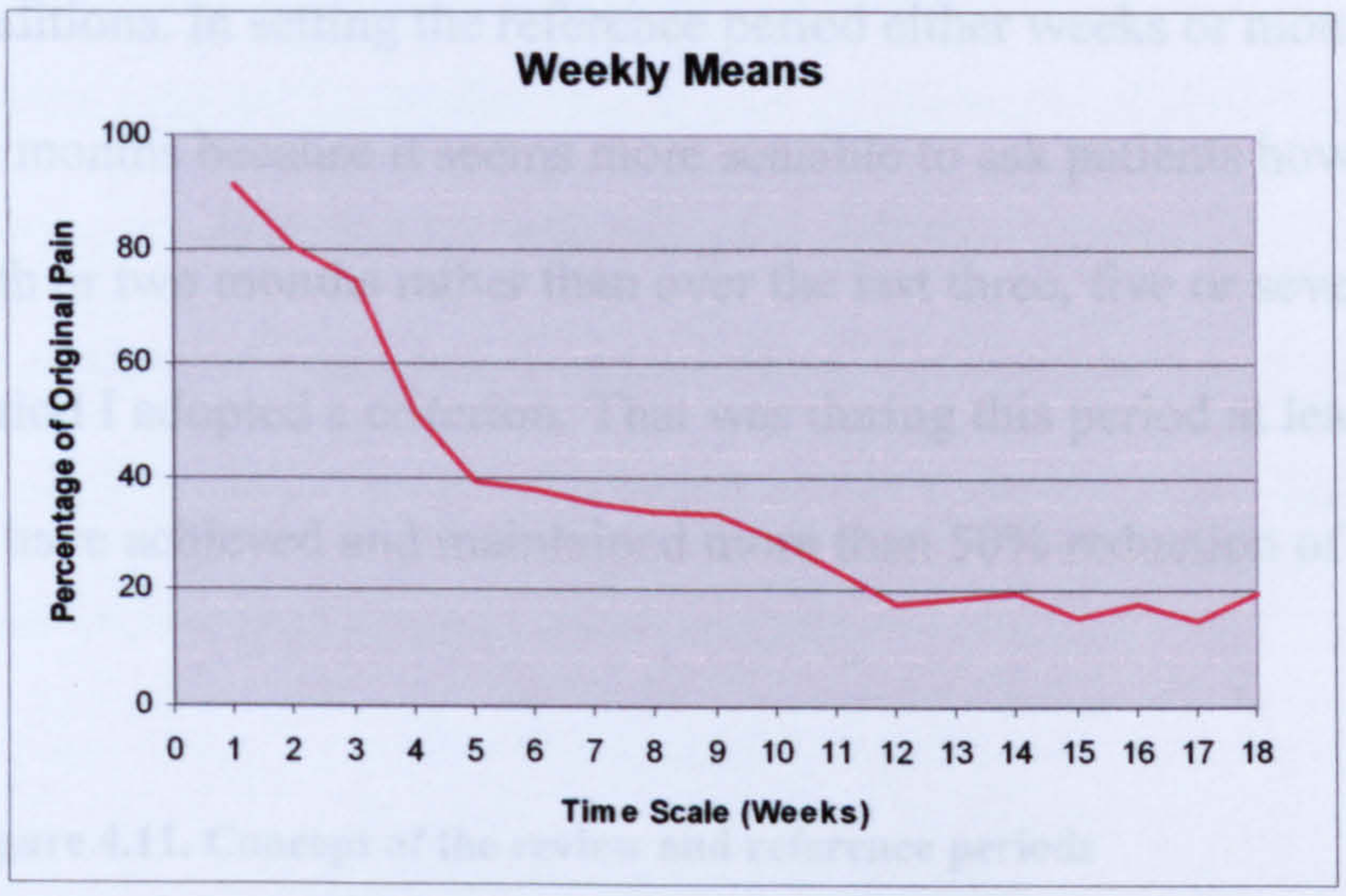
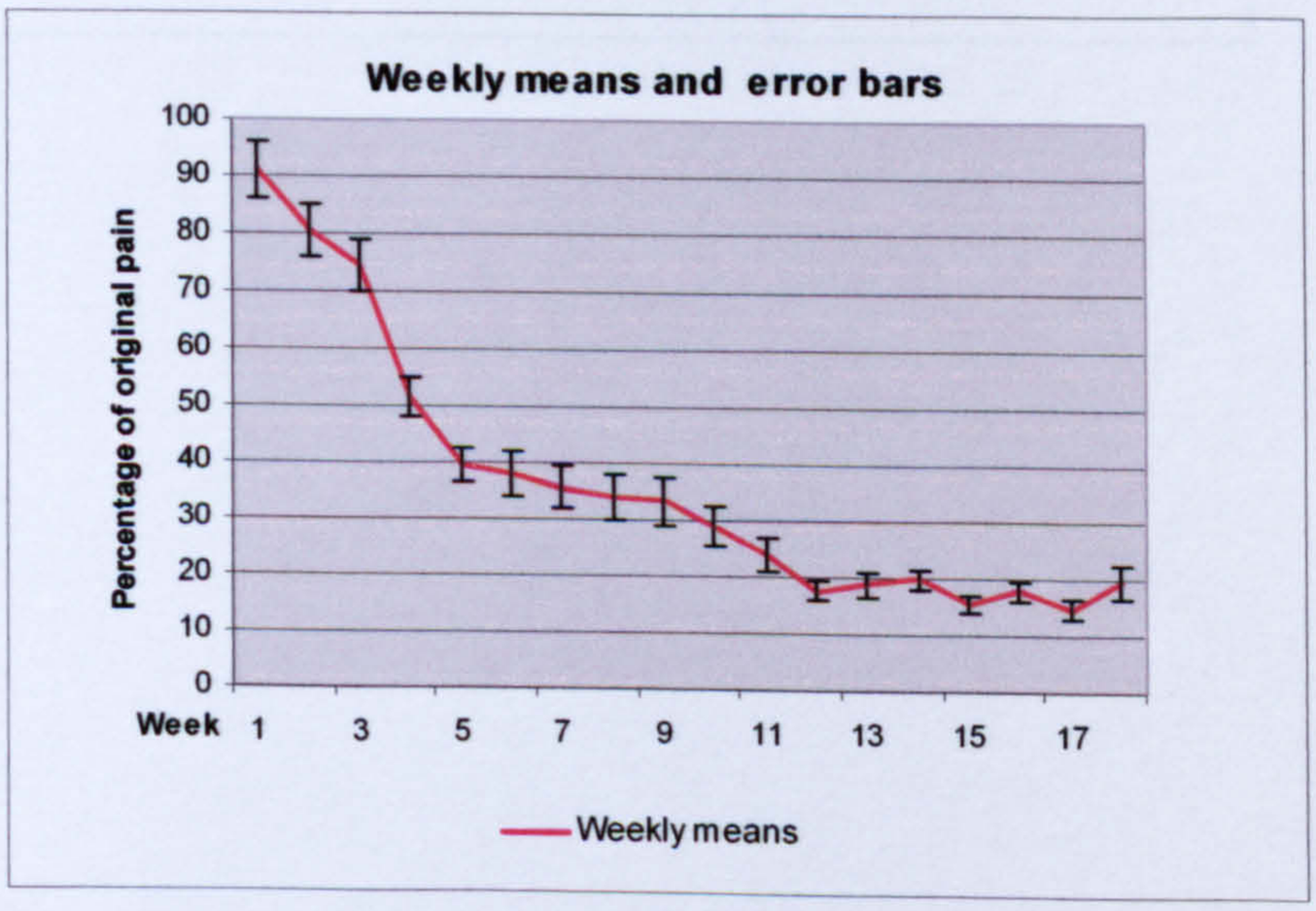
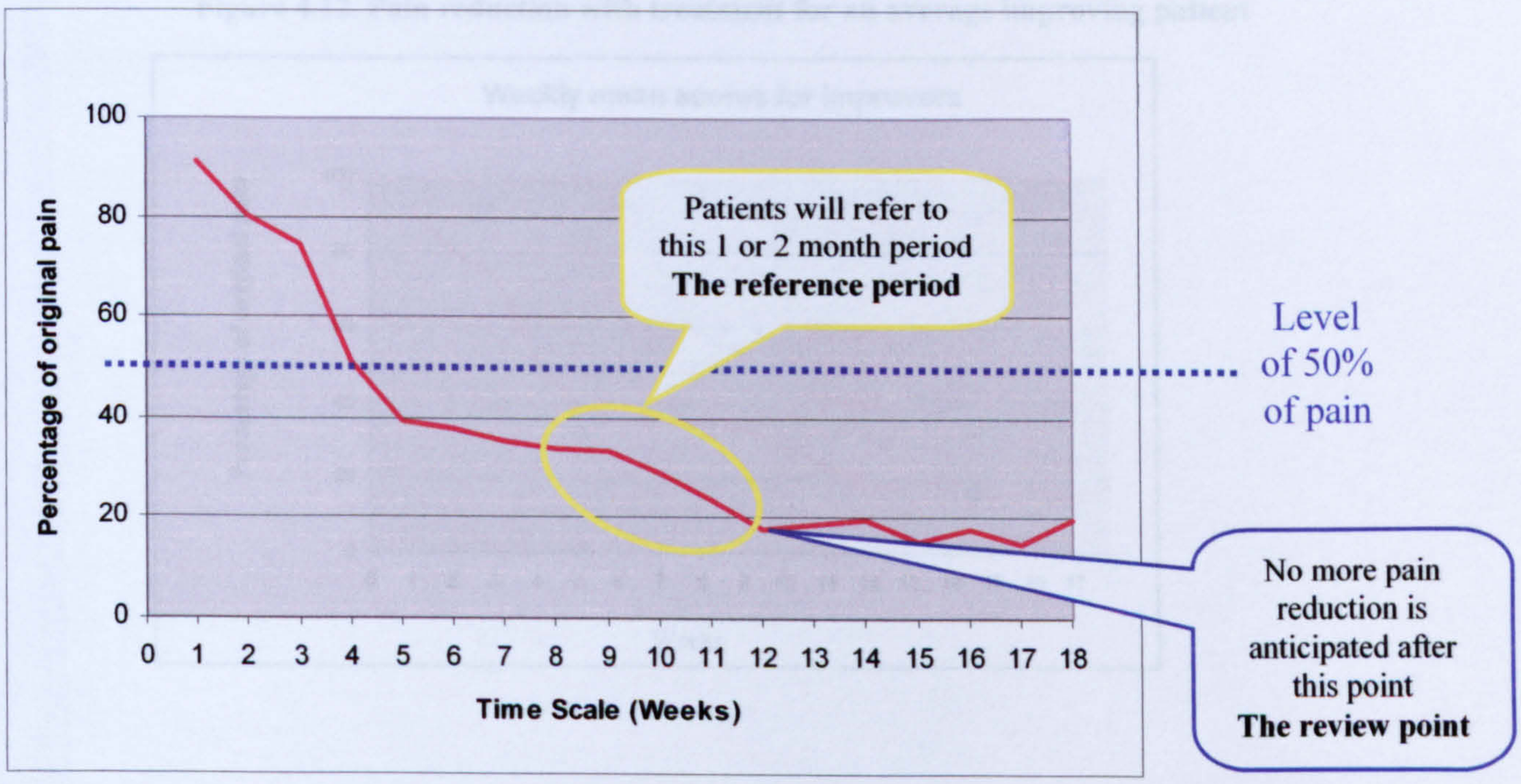


Figure 4.10. Weekly means and standard errors of means for improvers



The weekly graph was examined to determine the time at which improvement had leveled out, and this was chosen as the earliest possible review point (Figure 4.11). The reference period is the time before the review point over which the patients will be asked to appraise their conditions. In setting the reference period either weeks or months can be used. However, I chose months because it seems more sensible to ask patients how they have been over the last month or two months rather than over the last three, five or seven weeks. To determine this period I adopted a criterion. That was during this period at least 75% of improvers would have achieved and maintained more than 50% reduction of original pain.

Figure 4.11. Concept of the review and reference periods

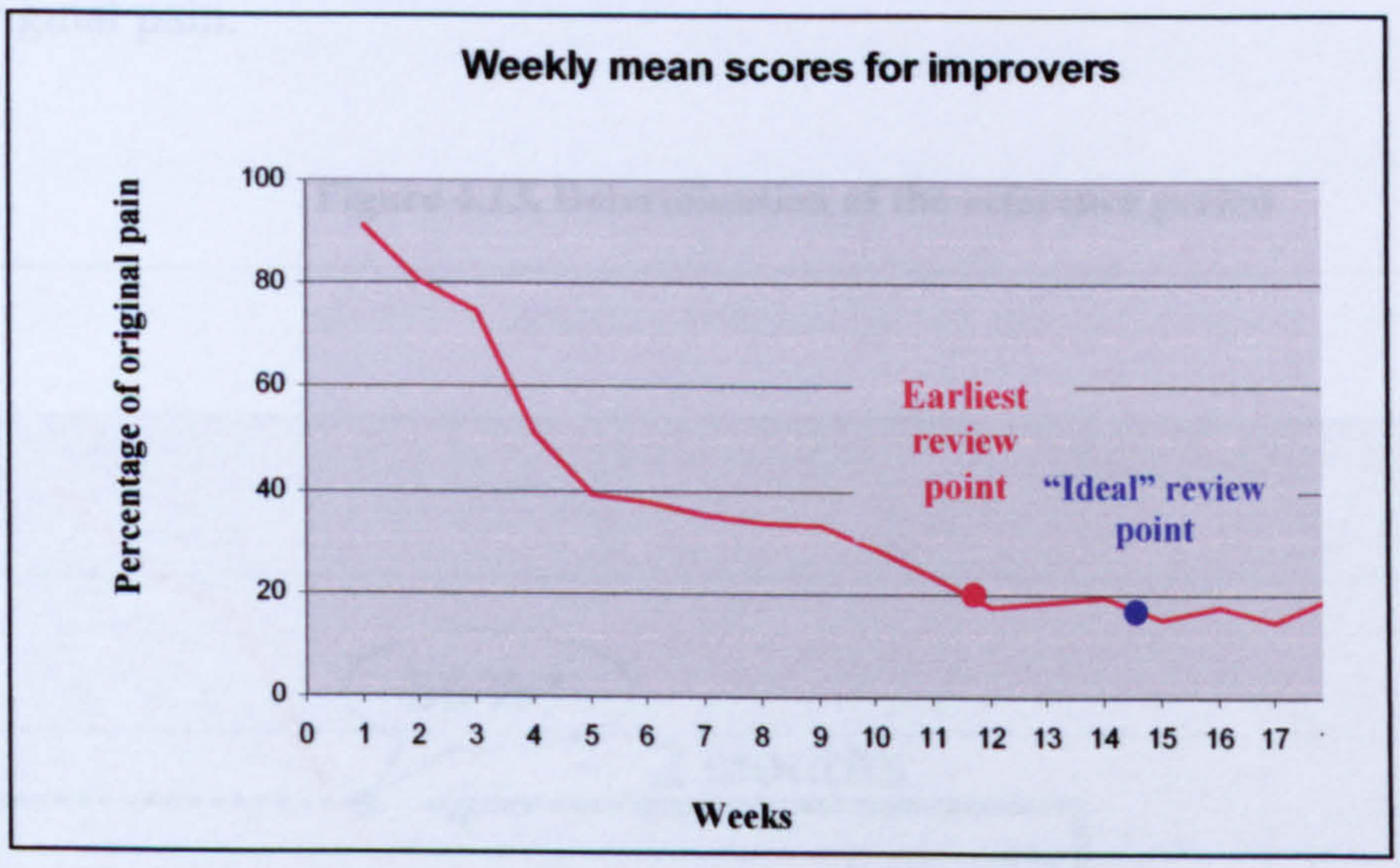


4.6.2. Results

4.6.2.1. The review period

The graph of the combined data from the 18 improvers (Figure 4.12) shows that mean pain reduced to 40% of the original level after five weeks of initiating treatment, including the period between taking the impressions and fitting the splints. Another 20% reduction occurred up to week 12. From then onwards, the mean pain level flattened around a value of just below 20% of original pain. I therefore defined 12 weeks as the "earliest review point". An "ideal review point" should allow extra time to maintain improvement.

Figure 4.12. Pain reduction with treatment for an average improving patient

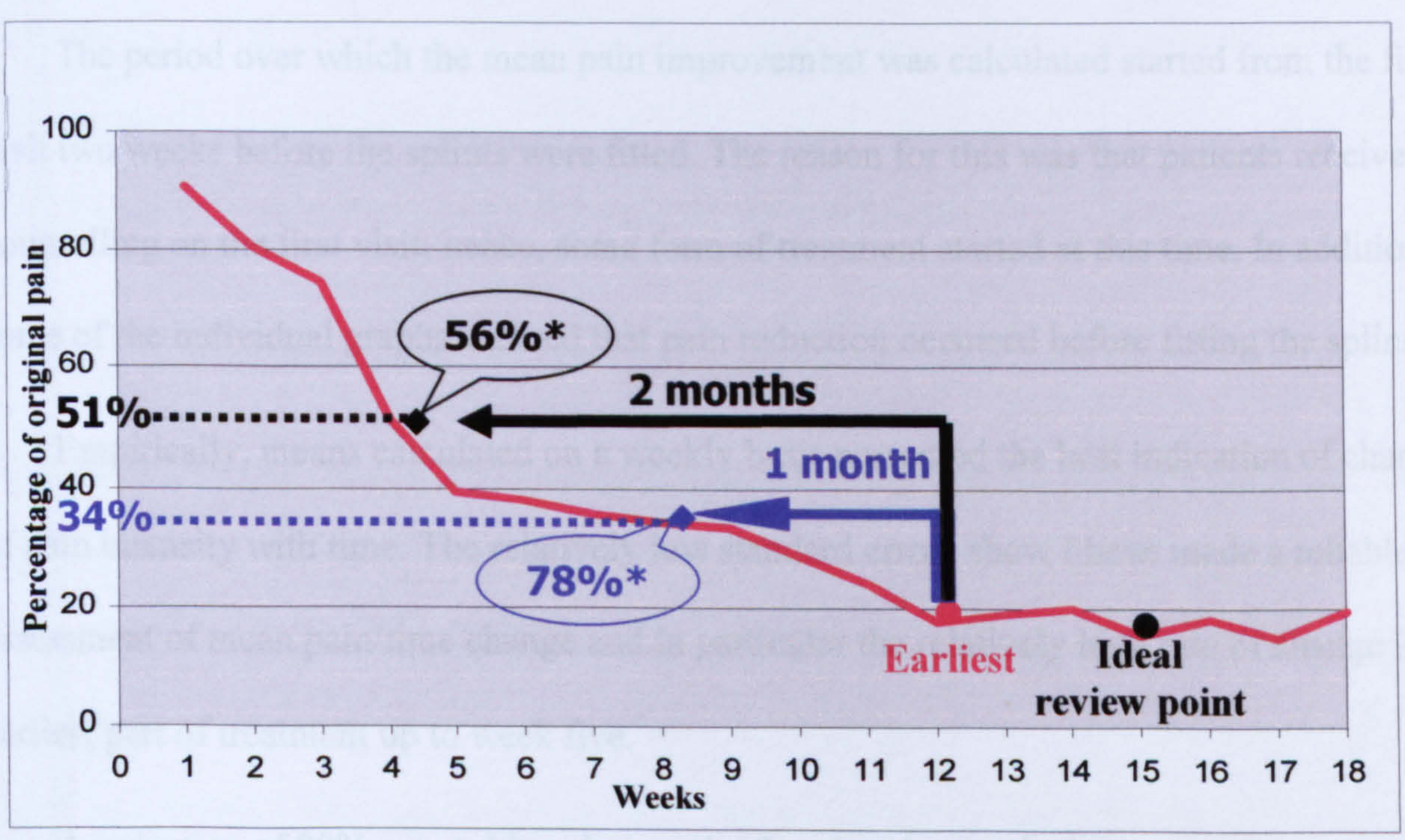


4.6.2.2. The reference period

To determine the effect of varying the reference period, two scenarios of one and two months were examined. When theses scenarios were applied from the earliest possible review point of 12 weeks (Figure 4.13), it was found that:

1. For a reference period of one month patients would consider their condition from the end of week eight onwards, a point at which almost 80% of improvers would have achieved ≥ 50% of pain reduction; and an average responder would have achieved 66% reduction in pain.
2. For a reference period of two months patients would consider their condition from end of week four onwards, at which point only 56% of all improvers would have achieved ≥ 50% of pain reduction; and an average responder would have achieved less than 50% reduction in the original pain.

Figure 4.13. Determination of the reference period



*** Percentage of improvers with ≥ 50% pain reduction**

4.6.3. Discussion

This study aimed at establishing optimum review and reference periods for TMD treatment outcome assessment. In the medical literature these periods have been chosen arbitrarily (see section 2.5). It was essential to determine these periods before developing and testing the intended quality of life instrument. For this, I examined the full course of pain changes for an average improver until improvement had stabilised.

I started by converting patients' scores to percentage of original score. This was necessary to adjust for the inter-subject variability in initial pain scores. It is also consistent with recommendations by other authors (Price *et al.*, 1983; Farrar, 2000).

As discussed previously in the development of the method, regression analysis gave results which did not made clinical sense. For that reason I excluded this approach and continued with the simpler method of calculating mean pain score for the improvers.

The period over which the mean pain improvement was calculated started from the first visit two weeks before the splints were fitted. The reason for this was that patients received counselling on the first visit; hence, some form of treatment started at this time. In addition, some of the individual graphs showed that pain reduction occurred before fitting the splints.

Empirically, means calculated on a weekly basis presented the best indication of change of pain intensity with time. The relatively low standard errors show I have made a reliable assessment of mean pain/time change and in particular the relatively high rate of change in the earliest part of treatment up to week five.

A reduction of 80% was achieved at week 12 and maintained afterwards. This suggests that the earliest review point should be after three months of treatment when, for the average responding patient, no more reduction of pain is expected. The latest review point should not be made much later than four months or it could compromise the initial management of non-

responding patients who may benefit from being seen sooner. The ideal point lies between the two limits. I chose 15 weeks, representing 3.5 months, as the ideal review period. Clearly, for a study of TMD outcomes it would have been impractical to have used either the six or twelve months reference period often used with OHIP.

To determine the optimum reference period I applied the two scenarios (i.e. one and two months) from the earliest possible review point (12 weeks). This is the most critical in setting the reference period because at the start of the period it would include more patients with a high level of pain than would a later review point (Figure 4.13). It was clear from this analysis that the two month-period's data do not meet the criterion previously defined, that is at least 75% of improvers would have achieved and maintained more than 50% reduction of original pain during this period. Furthermore, the retrospective assessment made with the two month-period will allow for a bigger recall bias than restricting this to one month (Redelmeier and Kahneman, 1996). Therefore, a reference period of one month was determined to be the best. This result is in agreement with the recommendations of the team who validated the German version of OHIP (John *et al.*, 2002). It is also in agreement with the reference period used in previous research using OHIP in TMD (Murray *et al.*, 1996; Segù *et al.*, 2005); where the choice of one month was based on the authors' clinical experience.

4.6.3.1. Strengths and limitations

This study is the first to suggest evidence-based review and reference periods to assess the outcome of TMD treatment. Deriving these periods was based on consideration of continuous data about pain changes throughout treatment as opposed to taking one or two "snapshots" of the full course of treatment. The study is limited, however, by the same factors reported for *Study 1-A*. Another limitation is the fact that I have considered here only improvement in pain. While this is the most reported symptom, other dimensions of the disorders should be assessed. The next study *(Study Two)* considers these domains. It is also important to mention that these results are limited to conservative TMD treatment, and that further work is needed in relation to other modalities of treatments.

4.6.4. Summary

I examined pain changes for patients reporting improvement with conservative treatment of TMD to establish suitable review and reference periods for outcome assessment. The assessment time for clinical trials should allow a clinically significant reduction in pain to be confirmed and maintained. For an average improver, a review time of 3.5 months was considered ideal. In addition, a one month reference period was more suitable than two months.

4.7. Conclusions of Study One

1. *(Objective 1.a)* – In chronic pain studies, different patterns of pain response to treatment can be encountered, representing improvers, non-improvers and including borderline cases. These distinct patterns could be used as an outcome measure on their own.
2. *(Objectives 1.b & 1.c)* – This work clearly shows that clinicians can make reliable qualitative assessment of chronic pain intensity. While this is a clinician-based interpretation of treatment outcome, it is potentially more representative than using a few snapshot measurements of pain and applying statistical tests blindly at intervals.
3. *(Objective 1.d)* – When using visual analogue scale (VAS) no single statistical method can be guaranteed to give a clinically valid outcome to determine improvers and non-improvers. The systematic approach described could be used to provide clinical validity to traditional statistical methods when measuring episodic chronic pain.
4. *(Objective 1.e)* – Study 1-A identified 20 improvers whose data were analysed in Study 1-B.
5. *(Objective 2.a)* – Assessment of TMD treatment outcome should allow sufficient time for improvement to be noted and maintained; it is recommended that initial outcome is measured at least three months after the start of treatment, and preferably not to exceed four months.
6. *(Objective 2.b)* – The reference period to be used at the review point should not exceed one month.

Chapter 5.
Study Two: Determination of suitable review and reference periods to assess TMD treatment outcomes:
A qualitative approach – Interviews with TMD patients

5.1. Introduction

The overall aim of this project was to develop a quality of life instrument to evaluate TMD treatment outcomes. One important consideration when designing such an outcome measure is the time at which it will be applied after the start of treatment, referred to here as "the *review period*"; a time which allows treatments to achieve outcomes noticeable by patients on their everyday lives. The other important consideration is the time frame of the individual items (questions); the *reference period.* While *Study One* analysed pain data to determine these two periods, *Study Two* had the same aim but used a different approach.

Qualitative in-depth interviews were undertaken in this study to explore perceived improvement and its time scale. This would allow for aspects of patients' suffering other than pain to be explored, which might be of significance to the sufferers. It would also allow for a triangulation of data from both quantitative and qualitative resources.

It was hypothesised that although pain is a major problem in TMD patients' lives, other symptoms may also play an important role. It is essential to take these symptoms into consideration when assessing the progress of the patient's condition with treatment, as different symptoms may require different times to resolve. In this context, interviewing TMD patients who are responding to treatment will offer a broader understanding of their experience of symptoms and perception of improvement with treatment compared to a series of visual analogue scale scores.

Although qualitative interviews have been used to develop measures of the psychosocial impact of oral health on quality of life such as the Oral Health Impact Profile (OHIP) (Locker and Slade, 1993; John *et al.*, 2002) and the Manchester orofacial pain disability scale (Aggarwal *et al.*, 2005), no study has explored the lived experience of TMD and its resolution, especially in relation to time. Only a few studies have considered patients' own accounts of the impact of other forms of oral pain such as toothache and burning mouth on

quality of life (Pau *et al.*, 2000; Anderson and Thomas, 2003; Hakeberg *et al.*, 2003; Anderson, 2004). In the medical literature, however, there are many qualitative studies covering the experience of pain, and probably the most relevant to this research are those investigating the lived experience of fibromyalgia and recovery from fibromyalgia (Raymond and Brown, 2000; Menghoel and Heggen, 2004).

While the original aim of this work was to verify the review and reference periods previously identified in *Study one*, great care was taken to understand the lived experience, and this produced findings with implications well beyond its original purpose.

5.2. Aim

To use qualitative methods to explore the lived experience of TMD patients throughout treatment, specifically including the process of improvement. The results will be used to verify the review and reference periods previously identified in *Study One*.

5.3. Objectives

1. To use qualitative data to explore the rate of recovery from different TMD symptoms as experienced by the patients.

2. To use qualitative data to establish the specific time scale of the recovery process.

As the interviews progressed a third objective was added to reflect both the need for additional data to understand recovery, and to explore the emerging themes. This objective was:

3. To explore the patients' journey through the process of suffering and treatment.

5.4. Methods

5.4.1. Subjects

The study used a purposive sampling method. Of twelve patients approached, ten were interviewed including nine females and one male reflecting the gender ratio for the condition (Table 5.1). They presented with different categories of chronic TMD according to RDC Axis I (Dworkin *et al.*, 2002a)[1]. All had been treated in the Restorative Department of the Newcastle Dental Hospital (NDH) by a combination of conservative treatment approaches. These included counselling, remedial jaw exercises, soft and hard occlusal splints, and in two cases, occlusal adjustment. Patients were included in the study if at review, after a minimum of three months of treatment they reported having less than 50% of pain in comparison with the start of treatment. Overall, 10 patients who met the inclusion criteria and agreed to participate were interviewed.

Table 5.1. Characteristics of the patient sample in *Study Two*

Number	Gender	Age	Diagnosis	Last treatment [4]	Treatment period [4]
1	F	65	MFP[2], DDWR[3], Arthralgia	Glucosamine, stabilizing splint	6 months
2	F	35	MFP, Osteoarthritis	Exercises, soft & stabilizing splints, physiotherapy	12 months
3	F	59	Osteoarthritis	Exercises, soft splint, analgesics	3 months
4	F	37	DDWR, Arthralgia	Exercises, soft & stabilizing splints, occlusal adjustment	4 months
5	F	52	MFP, Arthralgia	Exercises	9 months
6	F	54	MFP, DDWR	Exercises, soft splint	4 months
7	F	34	DDWR	Exercises, NSAIDs	3 months
8	F	47	MFP	Exercises, soft splint	4 months
9	F	33	MFP	Exercises, hard splint, Amitriptyline, occlusal adjustment	12 months
10	M	38	MFP	Exercises, stabilizing splint	4 months

[2] MFP: Myofascial pain
[3] DDWR: Disc displacement with reduction
[4] Last treatment in Newcastle Dental Hospital (NDH)

[1] A summary of the RDC criteria and the examination form used in this project are in Appendix 1

5.4.2. Preparatory stage

This study was conducted in collaboration with Dr. Catherine Exley, Centre for Health Services Research, Newcastle University. Advice was sought during the planning and analysis stages on detailed methodological considerations.

Approvals were sought from both Newcastle NHS Trust and the Local Research Ethics Committee LREC (Newcastle and North Tyneside). The study was registered in LREC under Ref. 2003/38.

Before taking part in the study, all participants were provided with a specifically designed information sheet and gave informed consent (see Appendix 4). The consultant in charge of the clinic was always informed about the interview to ensure a clinical reference point should any clinical issues be raised during the interview that required further investigation.

5.4.3. The interviews

A series of semi-structured interviews was conducted until all expected themes were explored and the full range of experiences had been investigated (10 participants). The interviews were conducted by the researcher and an experienced interviewer, Mrs Jill Smith using a pre-designed topic guide (Appendix 5) to define the broad themes to be considered during the interview. In order to avoid any bias in the data, for example, *"auspices bias"*, seen when patients tend to report higher satisfaction with treatment to avoid upsetting the treating doctors, the interviewers had not been involved in the treatment of any participant, and the interviews were undertaken in a quiet office within the Dental Hospital away from the clinical environment.

On average, the interviews lasted about 25 minutes, and were recorded using an audio digital recorder, backed-up with a traditional tape recorder. The audio digital files were

transcribed verbatim by a specialist company *(TypeFast 2000)*[1]. Digital record files and the magnetic tapes will be destroyed on completion of this higher degree.

Records' transcription and preliminary analysis were conducted immediately after each interview, which helped to identify emerging themes and to feedback and modify the topic guide. Only minor amendments were deemed necessary. For example, two sentences were added to the guide to encourage participants to give more details about their perception of improvement and about the time-line.

5.4.4. Data analysis

Data for this study consisted of direct quotations abstracted from the transcripts. Ideas emerging during the interviews were arranged within themes and were supported by the relevant quotes (Appendix 6).

The method used to abstract and analyse data in this study was the *Framework* method. It involves five separate but highly interconnected stages summarised below (Bryman and Burgess, 1994; Gantley *et al.*, 1999).

1. **Familiarisation.** At this initial stage the analyst familiarises him/herself with the data through reading and re-reading the transcripts. Initial listing of key issues and themes emerging from the transcripts takes part at this stage.
2. **Identifying a thematic framework.** At this stage the analyst builds a framework or an index based on the identified key issues and themes. This index is developed and refined using data from a few transcripts, and is designed to store data emerging from the remaining interviews.
3. **Indexing.** The developed index is systemically applied to each transcript to abstract all relevant data.

---

[1] Horseshoes, Cobgate, Whaplode, Nr Spalding, Lincolnshire PE12 6TD, info@typefast2000.co.uk , http://www.typefast2000.co.uk .

4. **Charting.** This is the stage at which data are combined from all interviews in one chart or matrix. This chart is devised from the previously established framework. It allows making comparisons between different respondents within the same "heading", i.e. theme; and also allows reading a summary of one respondent's data across all themes.
5. **Mapping.** Here pieces of the data set can be mapped together and analysed as a whole in order to make general interpretations and to address the *a priori* research objectives.

Data abstraction and analysis were conducted concurrently by two independent researchers, the author and one of the supervisors (JS), who then collaborated to develop a common thematic framework for final analysis.

5.5. Results

Four themes were expected from the topic guide and another one emerged from the interviews (Table 5.2). The first one, "length of suffering before seeking treatment and main reasons for seeking treatment", was a direct result of the topic guide, included to give some context to the subsequent interview, but was not considered as a core theme for this project. Detailed data from all five themes are presented in Appendix 6, but only themes b to e are presented here.

Table 5.2. Expected and emerging themes obtained from the interviews

Expected themes based on the topic guide
a. Length of suffering period before seeking treatment and main reasons for seeking treatment
b. Range of symptoms
c. Impact of TMD on quality of life (QOL) including functional and psychosocial disability
d. Perceived improvement and temporal characteristics of improvement
Emerging themes
e. Perception of care (primary and secondary)

In the topic guide, perceived improvement and its temporal characteristics were explored as one theme separately from other themes about suffering. However, for a better presentation of the results, improvement of symptoms and quality of life are presented along with the accounts of their corresponding sections. This is followed by a separate section on the temporal characteristics of improvement.

5.5.1. Symptoms of TMD and their improvement

This section describes TMD symptoms both in terms of their significance to participants and their characteristics. This is followed by an outline of symptom improvement.

5.5.1.1. Range of symptoms

While participants were not asked specifically to rank their symptoms in order, the order of the list of symptoms they reported was noted. A broad range of symptoms was reported. However, participants appeared to suffer most from orofacial pain and headache, and least from clicking/locking.

1- Pain and headache

All participants in the study reported facial or jaw pain. The characteristics of pain, however, varied noticeably. It was mostly, but not always, the first complaint reported. Headache was also one of the chief complaints, and in some cases the first to be reported.

The severity of pain and headaches ranged considerably amongst the interviewees, and with it the impact on quality of life differed. The pain on some occasions reached extreme levels:

> *"I was in just so much pain; I didn't want to get out of bed...it's the worst pain you can ever experience and I had two children and it's just the worst pain I could ever experience in me life, it was awful. Constant pain".* (Participant 2, Female)

> *"In fact I had a permanent headache and my face was feeling as if it was going to split open with the pressure. Because it was the pressure on the eyes and it was really, really painful, that was what was the worst".* (P5, F)

> *"The pain was to a point where it started being all down one side of my face to a point where the pain as that bad that my face was like paralysed... It's a pain that I never knew could exist. The type of pain even where there's times that I wished I was dead because the pain was just so unbearable".* (P9, F)

The interviewees gave specific descriptions about their symptoms, and while it was not the aim of this study to investigate this issue, these data were recorded and may potentially be used in future analyses. Table 5.3 summarises the range of locations and pain characteristics.

The periodical and fluctuating nature of TMD, especially the associated pain, was highlighted by many participants. Some even used to ignore the condition when they were having no pain from it:

> *"The pain changes from time to time...it was having relapses where the pain would ease. It would ease down to maybe 4-5* (out of 10) *and then within a couple of days it would shoot back up to about 10". "The pain can vary daily or... it can actually vary throughout the day.* (P9, F)

"Probably just discomfort because I find when I'm at work it's fine. Because I'm working, it's off your mind... probably because you're working it doesn't seem to bother you". (P10, M)

The need for taking medication including painkillers and antidepressants, was used by participants as a measure of the condition's severity as well as of the treatment success. The majority of participants needed painkillers on a different frequency according to their pain.

"I had morphine patches and pethidine and ... which weren't touching the pain... I was on a lot of medication for my complaints, what I've been through, but nothing, nothing took the pain away. Nothing at all". (P9, F)

"I'd take 8 a day when it was very painful...(but now) *not half as much, not hardly any actually".* (P2, F)

Table 5.3. Participants' own accounts of jaw pain and headaches

Location	Difficult to pinpoint or to localise Pain, ache, or toothache in the jaw Cheeks were sore In the side of the face/mouth Pain from the ears, pressure on the eyes Pressure in/round the face Radiating towards the neck and the arms	
Characters	<u>The jaw pain:</u> Unpleasant ache/discomfort Aching jaw Jaw feels sore Jaw is painful Toothache Pulling sensation Awful, agonising, horrible Shooting pain when it happens Excruciating pain The worst pain you can ever experience Festering sensation Debilitating pain	<u>The headaches:</u> Bad heads Dull headaches Pressure
Periodicity	Intermittent Constant Constant but worse some times Periodical	

2. Non painful joint-related symptoms

In general, joint-clicking seems not to have an important impact on the quality of life. Five participants reported having clicking in their jaw joints. However, they reported it low in the list of their symptoms, at least third following pain and headache. Only one participant complained of clicking before she mentioned other symptoms. She, and some other participants, could not, however, distinguish between clicking and locking. In other words, it was not clear whether they were describing Disc Displacement without Reduction, or Disc Displacement with Reduction. Similar to the pain symptoms, participants gave specific descriptions to joint-related symptoms, and these are summarised in Table 5.4.

"My jaw was clicking a lot, it felt like it was locking". (P7, F)

"I feel that it ... as if your bones are all being breaking". (P5, F)

"When I opened my mouth there was a tightness on the jaw and I used to feel that if I didn't manoeuvre it, it just cracked". (P1, F)

"Kept locking on the one side, on the right hand side"... (and eventually) *"It made its way back over the.. to same position"*. (P2, F)

Table 5.4. Participants' own accounts of non painful joint-related symptoms

Clicking	Cracking Funny noise Tightness on the jaw
Locking	Catching Participant needs to "manoeuvre" or to "wobble" the jaw to avoid it catching
Crepitus	Grinding Jarring Breaking bones

3. Muscle-related symptoms

While non-painful joint-related symptoms were of slight impact on quality of life and did not occupy much of the interview, muscle-related symptoms were numerous, reported more often and had a greater impact on everyday life. These are summarised in Table 5.5.

Table 5.5. Participants' own accounts of muscle-related symptoms

Tiredness and weakness	*"I used to wake up.. I felt tired and my mouth was tired"* (P1, F) *" The muscles would be tired"* (P4, F) *"That side's not the same as that side. .just weaker or something"* (P6, F) *"Jaw felt so heavy and tired"* (P4, F)
Paralysis and stroke appearance	*"You can imagine when people are having strokes their facial muscles all go, well that's what it felt like"* (P8, F) *"My face was like paralysed... I felt I couldn't move it... I looked as though I'd had a stroke"* (P9, F)
Stiffness and tightness	*"Some mornings I would get up and my jaw was so stiff it took me ages to get it working* *" My jaw felt tight" "There was a tightness on the jaw"* *" The muscles here always felt really tense"* *"The muscles running down the front of me face always felt like they were strained"*
Clenching and grinding	*"Sometimes when I woke I would be aware that I had ... my jaw was clamping really tightly shut".* (P4, F)

In addition to stiffness and weakness of the muscles resulting from hyperactivity or parafunction, one particular problem seemed to have additional impact on participants, and that was grinding the teeth and creating an aesthetic problem (see section 5.5.2.1).

5.5.1.2. Symptom improvement with treatment

One of this study's objectives was to explore how symptoms improved and how long each symptom required to improve. Reported here are characteristics of improvement. The time scale of improvement is presented later in section 5.5.3.

It was the reduction of pain and headache that participants reported as most important, consistently reporting this first and talking about improvement predominantly in terms of pain reduction. The pain was reported generally to have improved in terms of intensity, durability and frequency. For example one participant told of her experience:

"Before I came here if my jaw locked I would maybe have say 2-3 weeks of like toothache and pain in my jaw. But once they explained about the exercise I would say maybe about 3-4 days down the line I would start to feel... (Before the treatment the problem) (It) *would go away and it would be okay for a month. But once they showed me the exercises... I would maybe go maybe a couple of months again where I was fine".* (P7, F)

For some participants, pain has not been completely solved with treatment. Despite this, they found that the treatment helped them to cope better with their conditions than they did before:

"I think I'm coping a lot better than what I was coping with the pain". (P10, M)

"I wouldn't say I'm feeling better every day, but I'm feeling that what I've got I can cope with better". (P5, F)

The improvement of muscle-related symptoms was also important and was consistently linked with well-being more than improvement of the clicking and locking. This included having freedom of mouth opening, relaxation and more movement in the face, reduction of the stiffness, reduction of tooth grinding and clenching, feeling the jaw was secure and looking better:

"I've got more movement in my face...I feel better...My face feels more relaxed. There's been an improvement across my face". (P5, F)

"I think I felt happier plus I was sleeping better...You know, I wasn't' grinding me teeth as much as I was". (P10, M)

In parallel to symptom reduction, the participants considered the decrease in using the treatment an improvement on its own. This included taking fewer painkillers, decreasing the use of the occlusal splint and less need to visit the Dental Hospital for a follow-up.

"Very much improvement... I don't have the pain at all and my jaw doesn't click, you know, I can open my mouth freely, you know, and I don't live on painkillers". (P1, F)

"I've managed to go six or seven weeks without having to come up (to the hospital)... *which to me is, a move forward".* (P9, F)

5.5.2. The impact of TMD on quality of life and its improvement

5.5.2.1. The physical impact

The physical impact of TMD was far reaching, with oral function, everyday activities, work and sleeping all being affected.

The oral dysfunction involved eating, talking, yawning, and even smiling. Almost all participants experienced problems with eating, but this ranged from avoiding certain types of food to having extreme difficulty in taking any sort of food. Two types of food were particularly mentioned by participants; those that need a lot of time to chew such as meat and hard foods such as apples and toast:

> *"I couldn't eat nothing. I had to eat soup and drink pop all day ...*
> *I couldn't hardly open me mouth"* (P2, F)

> *"I used to have to just pick, you know, pick certain foods that I didn't need to chew because my jaw used to click quite a bit...* (But) *it just wasn't worth the hassle, you know, it was an easy option, not to eat meat".* (P1, F)

> *"Well I'm aware of what I eat.. Opening my jaw for an apple or something is a bit, you know, If I want to get a bigger bite".* (P6, F)

Difficulties were also encountered by some participants when opening the mouth during yawning, to receive dental treatment or even just to talk. One lady, a switchboard receptionist, suffered to the extent that she asked to be moved to another job:

> *"I worked as a receptionist and I found that if I'd been talking a long time, speaking a long time on the telephone it was really painful by the end of the evening ...and I actually asked if I could move off the switchboard. I was that concerned about having to speak for a long period of time".* (P8, F)

> *"I found when I had to go to the dentist ... she couldn't get in to do any treatment. She couldn't get in me jaw".* (P10, M)

Tooth grinding with resulting aesthetic problems were also experienced by one participant. She was aware of a link between the grinding and TMD symptoms, but was most distressed with the lack of confidence in her smile:

> *"I used to have nice teeth and the deterioration from that was drastic. When I looked in the mirror it sometimes doesn't reflect me. It changed my appearance so much. My teeth were changing and it was changing my whole face really... and my appearance".* (P1, F)

The impact of TMD on everyday activity was markedly different amongst the interviewees. For the majority, it did not affect their daily life, but for others there was a dramatic affect:

> *"Well it didn't stop me doing anything... I still was working".* (P3, F)

> *"No it wasn't affecting my daily life".* (P6, F)

> *"I just wanted to lie in bed all the time I was in just so much pain; I didn't want to get out of bed".* (P2, F)

One lady (P9), who appeared to be the most suffering person in the sample because of her long experience of unsuccessful multi-referral, reported the consequences of her pains:

> *"It stopped me doing everything. I've been off work for over a year"*
> *"I couldn't even do shopping. I was confined to the house really"*
> *"I was having to live with me mum and dad because the pain was that bad I couldn't look after my little boy. I couldn't look after myself".* (P9, F)

In addition to the daily suffering, almost all participants experienced sleep difficulties ranging from sleep disturbance to waking up at night or in the morning with pain or headaches:

> *"I've always been a good riser on the morning but I was very sluggish, you know, I would wake up with this headache".* (P1, F)

"It did affect you sleeping, I think more difficult really - It wasn't all the time- but it wasn't anything I could say I'd be worrying about". (P3, F)

"Sleep was impossible. The pain was so intense. I think the only time I went to sleep was when I think my body was just so run down and knackered with the pain I could sort of cat nap". (P9, F)

"I can't really sleep on this side because it's just an ache. So I ... I tend to sleep on that side anyway". (P10, M)

5.5.2.2. The psychosocial impact

The psychosocial impact of TMD varied considerably across the participants, not only in presentation but in severity as well. For some, it was a lack of self-confidence, which was linked mainly to pain but also to the appearance in one participant. For others, it was mood alteration such as depression, desire to stay alone, distress, anger, and despair.

"Well you feel a lot more confident in yourself when you're not in pain". (P2, F)

"(It affected) *my confidence because I had what ... well I couldn't smile with grinding my teeth the smile is only a part of it, you know, and when I look at photographs of when I was younger, there's no comparison".* (P1, F)

"I would say it would alter my mood when the pain was there". (P7, F)

"It was really distressing...It did make you feel quite low because it would happen in periods. So it was certainly affecting me in that respect (everyday life) *and yeah making me feel really just generally unhappy".* (P4, F)

"I used to cry with the pain, it got that bad...it was awful - Well the pain plus the worry of wondering what was wrong with me. That really got me down...I suffered with depression with it". (P5, F)

"Despite having fantastic little boy I just wished I was dead because I can't cope with this intensity of pain any more". (P9, F)

5.5.2.3. Improvement to quality of life with treatment

With the symptom improvement, participants reported an improvement to their quality of life. However, while there was a clear improvement to the psychosocial impact, the physical impact seems to be less improved, and the interviewees reported a long term functional disability.

For example, while the participants reported having better sleeps with treatment and waking up in the morning feeling more active with less or no pain, headaches or stiffness, most of them explained that they would still prefer to avoid hard foods and limit yawning even after achieving significant improvements to their symptoms and the psychosocial impact.

"It just became noticeable that I wasn't waking up with like really nasty pains or headaches". (P4, F)

"And I feel as though I can talk quite easily, you know, quite freely because I was a bit scared to speak". (P8, F)

"I still don't eat apples, that's something ... I'm frightened ... if I do eat apples I'll cut them, I won't bite into them". (P7, F)

"Well I still don't eat solid like hard chewy or toffee or boiled". (P5, F)

"I've got my life back in control. I'm back to work...I go out with my friends... I take my little boy to the park... This sounds really dramatic but Dr Wassell has given me my life back". (P9, F)

As regards the psychosocial impact, participants reported improvement in mood and wellbeing, an increase in self confidence, and enjoyment of life especially after knowing what was wrong with them:

"I had x-rays and so forth and he (the consultant) *said what the problem ... there was obviously no real sort of like cancer or anything like that which I was thinking... I was relieved that it wasn't tumour growing...when you feel a lump in the side..."* (P8, F)

"A bit more confident that I'm not going to just feel like the headache's coming on at any minute". (P4, F)

"It improved with my well-being as well. I'd get a sort of glow"
"I can have a day feeling as if I'm on top of the world sort of thing". (P5, F)

"I can sit back on the couch now put me feet up and watch Eastenders. And I know that in the end the pain's there but I sit on the couch and think I couldn't do this before". (P9, F)

5.5.3. Time scale of improvement

Since one of the inclusion criteria in this study was having reduction of pain of at least 50% of original pain with treatment, all interviewees reported improvement, but with differences in definitions, timing and rate of improvement.

Reduction of different symptoms followed different rates but jaw pain and headaches were consistently the most important symptoms and the first to improve.

Start of improvement

In general there was always a delay, and improvement was not noticed straightaway at the start of treatment. It was difficult for participants to pinpoint when improvement started to occur. Some found it took a month to start noticing some recovery of pain or headaches.

"It took a month probably" (to notice any improvement) (P3, F)

The same interviewee reported that later in the treatment period she would only wear the splint when she felt she needed it, and that would give her an immediate (next day) relief:

"When I have had any ache in at all I've used the splint, and it has been, you know, much better the next day". (P3, F)

Another interviewee could not remember when she had started to notice improvement, but reported a complete relief of her pain within less than one month:

"It took about 2-3 weeks to actually clear up... to get rid of the ache completely". (P7, F)

Participants also reported sudden improvement after a change in their conditions or in the treatment plan. For example, one lady found improvement after wearing a hard splint for six months only when this was adjusted:

"The splint took 6 months to notice an improvement, just when it was adjusted". (P5, F)

Progress of improvement

In general improvement was gradual and constant, but in some cases was interrupted. For example, certain treatments provided some relief of symptoms for a period of time, followed by the improvement ceasing or even relapsing; this led the participants to require further care or a change in treatment plan. One participant captured the general pattern quite clearly:

"It's gradually improving; it just seems to be going on one level now apart from the odd thing...the treatment never got worse". (P3, F)

"I found that it ... either my jaw had changed shape or the clamping and grinding had altered the shape of the splint. It wasn't fitting properly... and I was again getting headaches more frequently and then the strange clamping thing started". (P4, F)

"It (the pain) *seems to go through days where I might have a good couple of weeks or a good odd day. And then all of a sudden whoosh. The pain will ... I'll have like, what I call a relapse".* (P9, F)

End of improvement

A complete dissolution of symptoms was hardly ever experienced by participants. As seen from the examples presented above, some symptoms, such as pain would ease while others would remain, such as difficulty in eating. Generally, it was difficult to pinpoint how long it

took each symptom to resolve and also to determine an end of improvement. This perhaps just reflects the chronicity and periodicity of the TMD condition. So for some participants a complete recovery would be achieved within weeks, while for others pain would remain and treatment would only help them to cope better with the condition:

> *"It took about 2-3 weeks to actually clear up... to get rid of the ache completely"*. (P7, F)

> *"I wouldn't say my problems have been completely solved; I live with them better some days than I do others"*. (P5, F)

5.5.4. Perception of care

Participants' perception of the care they received was not originally planned for in the topic guide. However, this theme emerged during the interviews as a key issue that was worthy of further careful exploration. The topic guide was therefore updated to ensure this subject was covered.

In the following sections this theme is presented under two headings: primary care and secondary care. The term primary care is used here to describe any care received before reaching the TMD specialist clinic at Newcastle Dental Hospital, and included both general medical and dental practices, but in some case could have included other specialists.

5.5.4.1. Perception of primary care

Almost none of the participants had a clear idea about their condition before they were seen at the Dental Hospital. For example, they did not know that the headaches they were experiencing may be part of their TMD problems. Also, one participant was treated by her doctor for ear problems, to discover later that her symptoms were caused by TMD.

Surprisingly, the majority of participants reported that their dentists referred them to the hospital without offering them any kind of primary treatment or even any explanation or reassurance about their conditions. For example one participant said:

> *"They just kept looking at me teeth, they didn't say anything,... injections and drilling into me teeth all the time" - "I think it was here at the Dental Hospital that* ...(they told me) *it might be something to do with your jaw. It was never me dentist or nothing".* (Participant 2, Female)

Another participant explained how her dentist avoided taking responsibility for her care:

> *"I used to complain to the dentist quite frequently and I was really, really bad and I told her.. and she said well it wasn't a problem from her".* (P5, F)

One of the participants, who suffered extreme pain to a degree that she wished she was dead, went through an unsuccessful multi-referral experience. This led her to suffer longer and to feel angry about the health professional:

> *" I've got a lot of anger still inside me ... about the people in the profession like who didn't believe that they'd done anything wrong, like him who threatened to wire me jaw together when he obviously didn't understand about the bite of the teeth and that me bite was all wrong.* (P9, F)

Participants generally highlighted their lack of knowledge which resulted in deep concerns and worry about the seriousness of the condition.

The worry and silent suffering before secondary care

Three of the participants believed they had cancers before they were seen in the Dental Hospital. One of them opted to hide her thoughts and worries from her husband for several months until she was reassured about the cause of her symptoms.

"I was really upset about it, frightened more than anything because I felt as though I wasn't able to speak properly...I couldn't close my mouth properly at all and I was quite frightened. I didn't say this to my husband". (P8, F)

The worry by itself was enough motivation for some participants to seek treatment even without severe symptoms. One participant said:

"I was starting to worry about it, I was imaging everything" "I think... that was the reason (for seeking treatment), *that was the finish I thought".* (P3, F)

With lack of clear information and genuine worries participants needed social support and others' understanding of their suffering, but did not always find it. One interviewee reported:

"I mean you talk to people and they don't understand what you've got. It is a very painful thing... So people just sit and look at you". (P5, F)

5.5.4.2. Perception of secondary care

Participants' experience in Newcastle Dental Hospital was generally positive. This has to be interpreted, however, by the fact that all of the participants had achieved at least 50% reduction in pain at the time of improvement.

The interviewees consistently reported their satisfaction with the consultation and explanation they received at their initial appointment. Having someone to listen to their suffering and informing them about their condition and the degree of its seriousness and commonality had contributed a lot of improvement they had achieved:

"I saw Mr Barclay and... he examined my mouth, obviously, and I had x-rays and so forth and he said what the problem ... there was obviously no real sort of like cancer or anything like that which I was thinking. I'd been to the hospital and I'd been to me own dentist and then here at the dentist and then ... but it's been fine since I've been to the Jaw Clinic... I feel very confident now since I've been here, a lot more confident now". "Once I had that x-ray and Mr Barclay said "nothing" I felt a lot better in myself... I was relieved that it wasn't tumour growing" (P8, F)

"They explained about the joints, how the joint worked and that, and some people have just this problem... It's the relief; because you get it in your mind that it's something ... It's not a worry because I know what it is now". (P7, F)

"It's nice to be somewhere where you feel the people can appreciate what you're going through because... like obviously people who, in the dental hospital obviously it's their area of work and it's just a relief to get to somebody... to understand what you're going through". "I think I had faith in Dr Wassell because I feel he was the only person who listened to what I was saying...". (P9, F)

5.6. Discussion

This study was designed to triangulate with *Study One* in this project to establish optimal review and reference periods to assess TMD treatment outcomes. The *Review Period* should allow for the symptoms to improve to a degree that is both detectable for sufferers and meaningful for clinicians monitoring the recovery. The *Reference Period* used in the outcome assessment should also enable sufficient time for the patient to make a valid assessment of the condition, but should not be too long so that it refers to the start of treatment where little or no improvement had been achieved.

In contrast to *Study One* which analysed pain data in a quantitative approach, *Study Two* aimed, using a qualitative approach, to investigate patients' perceived concept of improvement after TMD treatment and to explore the temporal characteristics of this improvement.

TMD is a complex and multidimensional condition that involves pain, physical and psychosocial impacts on quality of life (Murray *et al.*, 1996; Turk, 1997; Suvinen *et al.*, 2005). Therefore, assessing TMD and its recovery merely from a pain point of view might undermine the validity of the assessment. In this respect, triangulating data from both studies would help by identifying the wider picture of TMD and exploring the characteristics of improvement for both painful and non-painful symptoms.

Triangulating qualitative and quantitative methods has been strongly recommended by some authors who recognise that a judicious combination of these methods has an important role to play in health services research (Barbour, 1999). Qualitative research is an indispensable exploratory tool that supports quantitative studies in three ways: in exploring the process of illness, in understanding perceptions and beliefs and to generate a theory that explains a phenomenon (Gantley *et al.*, 1999). In contrast, quantitative methods do not explore phenomena but work to quantify and measure them, and to prove an established theory.

Qualitative research including interviews does not aim and does not permit for statistical representativeness (Fitzpatrick and Boulton, 1994; Mays and Pope, 1995). Therefore, the sample size is determined in a way that allows data saturation to occur. This is to say, the size should be sufficient to explore all expected and emerging themes during the interviews or to build a theory. Consequently, the number of the study participants is usually limited in comparison with quantitative studies, and the participants are usually determined on a purposive method that warrants a reasonable representation of opinions or experiences to be represented (Gantley *et al.*, 1999). In this context, the sample in this study was reasonable for the purpose of exploring the temporal characteristics of improvement. However, a full exploration of emerging themes requires a larger sample and a different approach.

As one would hope, this study resulted in different themes emerging. Some were expected as a function of the topic guide, but another emerged during the interviews. These are discussed below.

5.6.1. Symptoms of TMD and their improvement

As anticipated, the participants reported that TMD had a multidimensional impact on their quality of lives, which, in addition to suffering from symptoms, included physical and psychosocial disability. Although pain was the main source of suffering, the main reason for seeking treatment, and was often the first and most important noted improvement, other symptoms and dimensions of quality of life were also important.

On the one hand, the major significance of pain in patients' lives supports the approach in *Study One* which assessed recovery time based on pain. This would seem to be a reasonable approach. On the other hand, the assessment of improvement and its timing should take other symptoms into consideration.

The major contribution of pain has been recorded previously (Segù *et al.*, 2005) and is not at all surprising. Using a quality of life measure, the Oral Health Impact Profile (OHIP) these authors noted that OHIP scores of patients with non-painful joint-related symptoms were not significantly different from the normal population. Murray *et al.* (1996) found that patients were able to tolerate the functional manifestations of TMD but not its pain.

Muscle-related symptoms were considered by participants to have a considerable effect on their lives, on their feelings and even on their appearance. The improvement of these symptoms was also important and was consistently linked with well being. In contrast, non-painful joint-related symptoms such as clicking seem to have only a moderate impact on quality of life. This finding is in line with previous reports by Murray *et al.* (1996) who found that quality of life scores, measured using OHIP, were not affected in patients with non-painful clicking joints.

In addition to the above, the present study revealed that other factors should be taken into consideration when assessing the condition and treatment outcomes. These include pain characteristics and the use of treatment. Firstly, it is clear from patients' narratives that pain has different characteristics that need to be considered when assessing treatment outcome. A simple measure of severity for example may be insufficient to evaluate the condition. The fluctuating nature of TMD pain, for example, was emphasised by most participants, which has a major impact on the way pain recovery should be assessed and this supports clearly the findings from *Study One*. In addition, the nature of the pain, for example, being "dull", "shooting", "pulling sensation" or "festering" is also an important aspect of assessment as the nature of pain might affect the patient as much as the intensity itself.

Secondly, a reduction in the use of treatment was linked to symptom improvement and was considered an improvement on its own, where patients would take less medication and need to visit the hospital less often. This aspect of recovery should therefore be also considered when assessing treatment outcome.

5.6.2. The impact of TMD on quality of life and its improvement

It was also clear from the data that TMD resulted in a considerable physical disability, principally in terms of oral function, but not uncommonly in more general functioning and activity. This impact showed marked variation between the interviewees, both in terms of nature and severity. Also, improvement in these disabilities differed. For example, improvement to sleep made a significant contribution to quality of life, but avoiding certain foods seemed less important to patients and in fact did not respond to improvement. In other words, there was a long-lasting physical disability for some people.

The psychosocial disability such as lack of self confidence and mood alteration caused by TMD varied enormously between the interviewees. For some it was something that "can be put at the back of their mind", but for a minority of others it represented a complete stopping of normal life. Assessing this disability appears to be crucial to identify and manage the minority of sufferers with extreme QOL impact. Unlike the physical disability, the psychosocial impact has been effectively managed by treatment and in particular by providing satisfactory assurance and care.

In conclusion, TMD symptoms and their impact on quality of life appear to have unequal importance to sufferers. Pain in particular was the dominant symptom and the one to which improvement was principally linked. However, it is essential that treatment outcome assessment considers the different impacts of TMD on quality of life and allows sufficient times for them to resolve.

5.6.3. Time scale of improvement

As was expected, retrospective recall of time-related information was difficult for patients, and therefore data could only be considered as a general guidance. For example, the interviewees gave wide margins of time of changes in the condition, and participants tended to refer to months instead of weeks.

It was clear that different symptoms required different times to improve. In general, there was a delay in starting to notice an improvement, but it cannot be concluded whether this delay was caused by the time needed for symptoms to improve or by the failure of initial treatment.

The fastest improvement in the condition was noticed immediately after the first visit to the Dental Hospital, when patients were assured about their conditions. This "immediate" recovery might explain the sharp reduction of pain seen in the plotted graphs in *Study One* in the first few weeks after starting treatment, even before receiving a physical treatment such as an occlusal splint.

Apart from this initial swift change, improvement and its timing were mainly linked to reduction in pain, as discussed above. The participants reported gradual yet fluctuating improvement. The pain was also reported to have improved in terms of intensity, durability and frequency. Both pain and mood were reported to have changed from day to day and within the same day. This change in the condition is strongly linked to the results of *Study One.*

The implication of this is that the use of a short reference period in the assessment, for example, asking the patients about their pain at the day of assessment would subject the assessment to major bias by measuring a high or low level of pain only by chance.

Similarly, assessing treatment before reaching a reasonable plateau and stability of the condition risks obtaining a misleading picture about the condition where these symptoms could be high or low merely based on the day of assessment. The facts that improvement is noticed after a period of time, that it occurs gradually, and that it is fluctuating emphasise the

need to allow enough time before assessing treatment outcomes. Within the time frame reported in this study, the assessment should allow a few months for patients to notice a degree of recovery. This would also allow slow improving symptoms other than pain to be captured by an outcome measure. However, it should be noted, that not all symptoms should be expected to disappear, and some level of suffering may remain, such as long term functional impacts.

The findings about temporal characteristics of improvement in this study are consistent with the review and reference identified in *Study One*.

5.6.4. Perception of care

It was not surprising that different individuals with different symptoms, ages, general health and attitudes towards health would suffer differently. However, the psychological suffering and its extent before reaching a TMD care facility in the Dental Hospital were striking. Participants consistently described their lack of knowledge about their conditions before they arrived at the Dental Hospital. This by itself becomes a reason for suffering as some patients would start to worry silently about their conditions and imagine a range of scenarios. With these worries and a lack of satisfactory assurance and management by health personnel, patients would enter a vicious circle of agitation, worry, anger towards the health professionals, and more suffering.

The lack of diagnosis and the experience of multiple referrals had a major impact on quality of life. For some participants, these worries were even the main or the only motivation to seek medical advice although they were experiencing a wide range of symptoms. The implication of this finding would later present as an immediate relief when a clear and convincing communication was received from the specialist. This does not only highlight the importance of communication in the management of TMD, but also has an impact on the time scale of recovery as discussed in section 5.6.3.

5.6.5. Strengths and limitations

This was the first study to explore in-depth the lived experience of TMD. Also, this is the first work that endeavours to propose a suitable review time to assess outcome of TMD treatment. This study is limited, however, by a number of factors.

To start with, the small sample size does not allow generalisation of the results outside the context for which it was designed. This is the case for any qualitative study which is, by its nature, exploratory and cannot claim statistical representativeness.

Secondly, the emerging data about patient's suffering and how this related to early care were initially beyond the scope of this project, which in the fist instance was mainly concerned with the appropriate timing of assessment. Indeed, these data initiated an intensive piece of work taking place currently in the same research team, with the aim of developing a theoretical model to understand the patients' journey through illness and treatment, and to propose changes to the management strategy for TMD.

Thirdly, patients in this study were interviewed at one of their follow-up visits where complete recovery and improvement had not been achieved. This was necessary to explore the lived experience of improvement close to real time and avoid a long recall period which might introduce a memory bias. Although this is a potential limitation of the study, by not reporting long-term improvement, the aim was to establish the time at which initial improvement occurred allowing a meaningful treatment outcome assessment to be made. This is appropriate for clinical trials for example. So, in this respect, long-term monitoring of the participants was not relevant.

Finally, since the study sample was limited to patients showing improvement to treatment, data related to suffering cannot be generalised to all TMD patients, as those not benefiting from treatment might present different profile of suffering.

5.7. Conclusions

1. *(Objective 1)* – The impact of TMD on quality of life is multidimensional and includes suffering from pain and other symptoms, physical and psychosocial disability. Pain, however, was the main source of suffering, and its recovery was often the first and most important noted improvement. This finding supports the approach in Study One which assessed recovery time based on pain. This would seem to be a reasonable approach.

2. *(Objective 1)* – TMD different symptoms and impacts on quality of life follow different rates and characteristics of improvement. Timing of outcome measurement, should therefore take this into consideration.

3. *(Objective 2)* – Improvement of TMD is gradual and fluctuating. Therefore, outcome assessment should allow sufficient time for improvement to be noticed and maintained. This study concluded that the review period should be around a few months after start of treatment, which is in accordance with the conclusion of the first study.

4. *(Objective 2)* – Because of the daily fluctuation and gradual nature of TMD improvement, the reference period used in outcome assessment should be sufficiently long to allow a valid assessment of the condition to be made, but not too long so that it refers to the start of treatment where little or no improvement had occurred. In the light of the patient's accounts, a reference period of one month would seem reasonable.

5. *(Objective 3)* – The results of this study reinforced the importance of clear communication and reassurance as key tools in the management of TMD participants.

5.8. Summary

Study Two was conducted to triangulate with *Study One*. The aim of *Study Two* was to explore TMD patients' perception of improvement with treatment, to investigate rates of recovery from the various symptoms, and to verify the review and reference periods identified previously in *Study One*.

I used a qualitative approach, which involved in-depth interviews with patients undergoing conservative treatment for TMD and showing improvement. The interviewees consistently informed that amongst all TMD symptoms, pain had the most impact on their lives and was the symptom to which improvement was linked the most. This finding confirms that the approach adopted in *Study One*, i.e. determining the review and reference periods based on pain recovery was a reasonable approach.

The temporal characteristics of recovery reported by the interviewees, e.g. the gradual fluctuating improvement, are consistent with the findings from *Study One*. However, using a qualitative approach did not allow for an accurate determination of optimal review and reference periods. Estimation of these periods could be only conducted within large margins of time. Nevertheless, within these limitations time frames for the review and reference periods suggested in both studies are consistent.

Chapter 6.
Study Three: Establishing a quality of life measure of TMD treatment outcomes

6.1. Introduction

The impact of TMD on quality of life is well documented in the literature, though few attempts have been made to quantify its impact. RDC, the well-established and widely-used Research Diagnostic Criteria (Aaron *et al.*, 2004) does have a section (Axis II) to measure pain-related disability and the psychological status of the patient, but this does not take into consideration the social component of quality of life.

Health care providers are increasingly demanding proof for the benefit of the resource they provide. For them, their "customers' opinions" are as valuable as the professionals'. Bodies such as the National Institute for Health and Clinical Excellence (NICE) in the U.K., which determines the relative values of different treatments, are increasingly demanding that the impact of illness on the patient's quality of life is evaluated, allowing treatment decisions to be made taking into account the "units" of quality of life they can provide for patients (National Institute for Health and Clinical Excellence, 2006). The literature review discussed in detail the need for quality of life measures to be developed and used in health care assessment.

In order to allow this approach, appropriate outcome measures are required that can record, not only the impact of the condition on quality of life, but also the change in that impact with treatment. Various instruments are available to measure the psycho-social impact of dental problems. Most of these are concerned with oral health in general, and include the Dental Impact Profile (Strauss and Hunt, 1993), the Oral Health Impact Profile (OHIP) (Slade and Spencer, 1994) and the Oral Health Quality of Life – UK (McGrath and Bedi, 2001)[1].

Only one instrument was developed specifically targeted at orofacial pain, the Manchester Orofacial Pain Disability Scale (Aggarwal *et al.*, 2005). While promising, this has not been extensively tested. For example, its sensitivity to changes with treatment has not been determined. The instrument was established from a pool of patients with orofacial pain

[1] A more comprehensive review of oral-health related quality of life measures is presented in the literature review section 2.6.

including those consulting dental casualty and oral medicine clinics, but patients with non painful TMD conditions were not represented. So, it is specific to orofacial pain but not to TMD.

Amongst other oral health-related quality of life (OHrQOL) instruments, the Oral Health impact Profile (OHIP) (Aggarwal *et al.*, 2005) is perhaps the most widely tested and used. OHIP is a patient-based questionnaire that is structured in seven domains, inspired by the World Health Organisation (WHO)'s conception of impairment, disability and handicap (World Health Organization, 1980). A preliminary study by Murray *et al.* (1996) showed it to have the capacity to measure the associated impact on quality of life for TMD and orofacial pain patients (Aggarwal *et al.*, 2005). OHIP, however, has 49 items and covers all aspects of oral problems. Being such a comprehensive instrument, it is cumbersome and may cause respondent burden affecting completion rates. Moreover, in relation to TMD it contains many items that are potentially redundant, reducing its discriminatory ability and possibly making it appear of limited relevance to those affected. For these reasons, shorter condition-specific versions of OHIP have been produced, for example OHIP-EDENT for edentulous patients (Allen and Locker, 2002).

OHIP has been translated into many languages, and, as part of validating the translation, workers (John *et al.*, 2002; Larsson *et al.*, 2004; Segù *et al.*, 2005) have applied the instrument on TMD patients. The original English version, however, has not been examined in this context, nor has it been tested for its ability as an outcome measure for TMD treatments.

The aim of this study was to start the process of developing an outcome measure for TMD, using as a starting point an existing measure of oral health (OHIP), which has been shown to be sensitive to the impact of TMD. While mainly intended as an outcome measure, this instrument should also be able to discriminate between TMD patients and TMD-free subjects. The derivation process of the new tool starts by identifying OHIP items impacting most on TMD patients. This is followed by examining aspects of validity and reliability of OHIP-49 in an English-speaking TMD population, and exploring how these properties may change if redundant items were eliminated.

6.2. Aim

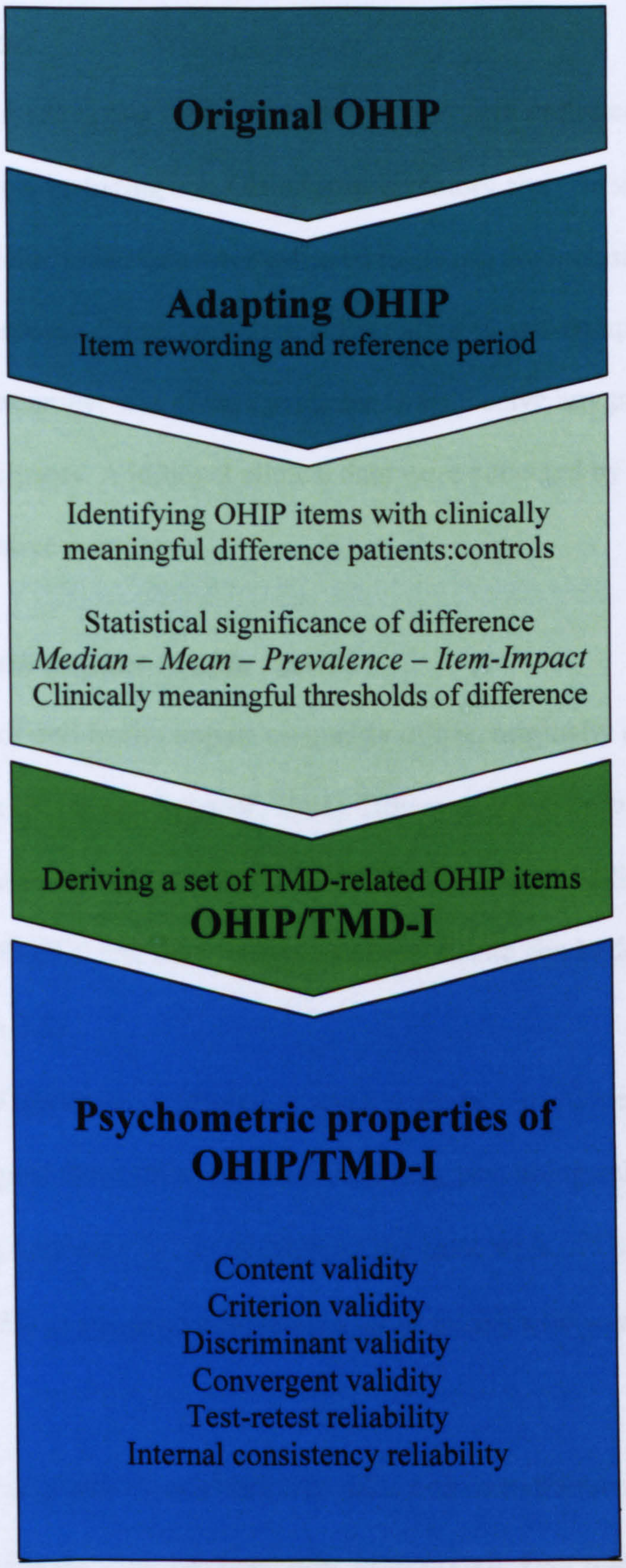
To begin the process of developing an outcome measure for TMD, using as a starting point the Oral Health Impact Profile (OHIP).

6.3. Objectives

1. To adapt OHIP-49 to make it meaningful to TMD patients
2. To collect full OHIP-49 data from patients of known TMD diagnosis (RDC/TMD- axis I) and from TMD free controls
3. To evaluate and compare the behaviour of each OHIP item in both TMD and controls by four different methods: *Mean, Median, Prevalence and Item-Impact*
4. To remove redundant items and select a subset of candidate items for a short form TMD index, the so called "OHIP/TMD-I"
5. To explore some psychometric properties of OHIP-49 and "OHIP/TMD-I" in an English speaking TMD population. These include aspects of validity: content, criterion and construct (discriminant and convergent) validity; as well as stability and internal consistency reliability.

An outline of the study is presented in Figure 6.1.

Figure 6.1. Outline of *Study Three*



6.4. Methods

6.4.1. Data collected

OHIP was the core instrument in this study. Data from OHIP were collected in parallel with other validating instruments including a validated pain inventory and two single-item scales of pain and perceived oral health. Other data were gathered to ensure the inclusion criteria were met and for follow-up purposes. These data were collected using self-completion questionnaires, but the researcher was always available to help solve any problems or questions raised by participants. Additional clinical data were collected by the researcher for diagnostic and administrative purposes.

6.4.1.1. The Oral Health Impact Profile (OHIP)

OHIP is a measurement of oral health impact on quality of life, originally developed for epidemiological research (Slade and Spencer, 1994). However, it has been used successfully as a measure of treatment outcomes (Awad *et al.*, 2000; Allen and McMillan, 2003). A more comprehensive review of OHIP and other measures of oral health can be found in the literature review (section 2.6).

OHIP consists of 49 questions structured in seven domains: functional limitation, physical pain, psychological discomfort, physical disability, psychological disability, social disability and handicap (Appendix 7). An example of the basic style of OHIP items is: *Have you found it uncomfortable to eat any foods because of problems with your teeth, mouth or dentures?*

OHIP uses a Likert response format with five units related to frequency of impact in the following order: *Never, Hardly ever, Occasionally, Fairly often, Very often*; which are coded 0-4. To improve the appropriateness of OHIP with TMD patients, I slightly modified the questions by adding the word *"jaws"* after *"because of problems with your teeth, mouth, or*

dentures". This ensured that participants were able to relate the questions to their TMD. The modified version of OHIP used in this project is presented in Appendix 7.

The reference period in the original OHIP is one year. In other words, patients are asked to assess their quality of life over the past 12 months. The first and second studies in this project established that the reference period for questions for TMD patients should be framed at one month. Therefore, the time reference of the OHIP in this study was one month.

To identify the TMD relevant questions in OHIP, it was administered to two groups: TMD patients and healthy controls. For patients the instrument was administered twice, pre- and post- treatment, with the aim of exploring its sensitivity to changes with several methods of TMD management.

6.4.1.2. The Multidimensional Pain Inventory (MPI)

In order to examine the convergent validity of OHIP in the TMD population, the working instrument had to be validated against another measure of TMD and its impact on quality of life. To that end, the Multidimensional Pain Inventory (MPI) was considered "the gold standard" multiaxial pain self-report instrument (Newton-John, 2002).

MPI (Kerns *et al.*, 1985) is a generic assessment tool for the impact of chronic pain on quality of life. It has been validated and extensively used in pain research. Further details are presented in section 2.2.6.1. The questionnaire comprises 52 items with a six-point scale (Appendix 8). These are divided into three groups: *Part 1* consists of 20 questions about pain's impact on the patient's life. *Part 2* has 14 items asking about the support the patient receives from spouse and close relatives. *Part 3* contains 18 items about the impact of pain on the patient's daily activities such as doing house-hold chores. Since parts two and three were of minor relevance to TMD sufferers, only part one was used. This part has some questions relating to the other parts, for example item two (pain interference with daily activities) covers

at least ten items in part three, and items five and ten (about the response of spouse and significant others), which cover most questions in part two. This questionnaire (MPI- part one) was administered only to the patient group, on two occasions pre- and post treatment. The second application was necessary to explore OHIP's sensitivity to change with treatment in future work.

6.4.1.3. The Visual Analogue Scale (VAS)

VAS scales of pain intensity were used to examine the construct validity of OHIP. Similar to the MPI, these scales were administered twice, pre- and post-treatment. On both occasions two VASs were used. The first (VAS1) was related to current jaw pain (pain at the moment), while the second (VAS2) asked about jaw pain over the last month.

The standard VAS was used, which was a horizontal line of 100mm long, bounded with "no pain" and "unbearable pain" as the narrative "anchors". A review of the literature on VAS is presented in section 2.4.4.4.

6.4.1.4. Global statement of health status of jaws and mouth

In addition to self-report measures of pain, a global statement of oral health pre and post-treatment was used for two reasons:

1. The first application was used to examine the convergent validity of OHIP and the new short form (OHIP/TMD-I)
2. The difference in this global rating between its initial and follow-up applications will be used to examine OHIP's sensitivity to change with treatment in future analyses.

The statement read: *Overall, would you say that the health of your jaws and mouth is: Excellent, Very good, Good, Fairly good, Poor?*

6.4.1.5. Other data

Other data were collected to confirm the inclusion criteria (see section 6.4.2.1), including:

- *Do you have any pain or problems with your jaw muscles or joints for which you need treatment?*
- *How long have you had jaw pain or problems for? ...Years ...Months.*

The following two questions were asked for follow-up purposes:

- *How effective do you feel the treatment was for your pain or jaw problems: Excellent, Very good, Good, Fairly good, Poor?*
- *Do you think you still need treatment for pain or problems with your jaw muscles or joints: Yes, No?*

All patients were examined clinically in a systematic way as described later. For every patient a summary of the diagnosis and treatment plan was recorded by the researcher. In addition, the full diagnostic form adopted in this research (Appendix 1) was completed whenever possible.

Finally, a pre-coded registration form was completed for all patients and controls to record names, age and gender. This information was kept separate from all other forms used to ensure anonymity.

6.4.2. Study subjects

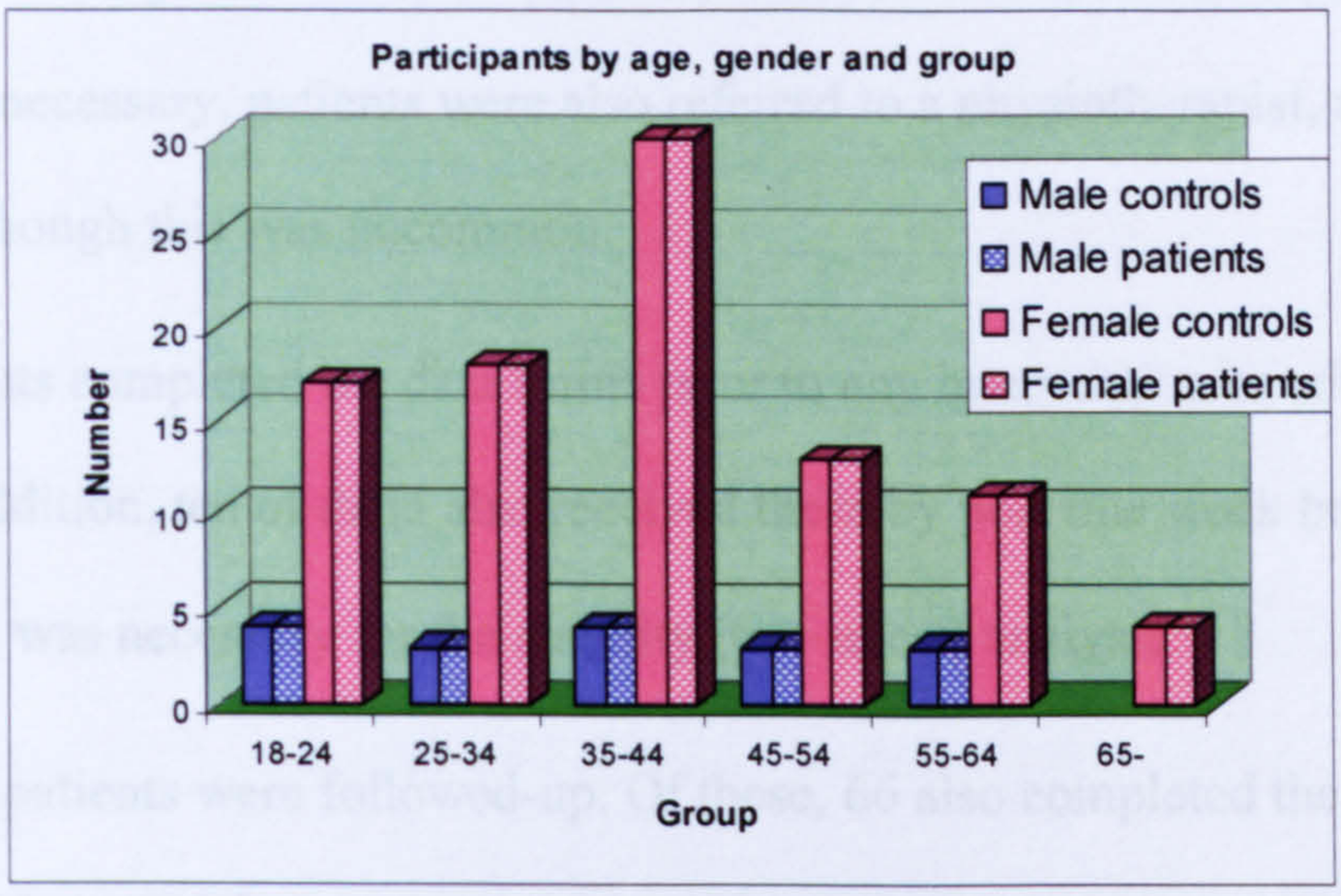
The project used a case-series design. TMD patients completed OHIP and other forms twice, before and after treatment, while the control group completed OHIP only once. As mentioned before, the post-treatment data is not presented in this thesis, and is to be analysed in future.

6.4.2.1. TMD patient group

Out of 126 patients initially invited to take part in the study 110 agreed to participate. These were newly referred TMD patients recruited consecutively from five restorative clinics in Newcastle Dental Hospital. Patients were from both sexes, but were predominantly females (84%). They were aged 18-82 years with a mean age of 39 years. For the purpose of matching

with controls, these were grouped into six age bands (Figure 6.2). There was no statistically significant difference in age, gender or diagnosis between those who participated in the study and those who did not.

Figure 6.2. Distribution of patients and controls in the different age and gender groups



Inclusion criteria

All patients were systematically diagnosed. They were included if they were 18 years or older, had symptoms for more than three months and been allocated one or more RDC/TMD diagnoses (Appendix 1). Patients presented a wide variety of chronic TMD conditions and sub-diagnoses (Table 6.1).

Table 6.1. Distribution of patients within different diagnostic groups
(More than one diagnosis is allowed)

Group I	Group II	Group III
Myofascial pain	**Disc Displacements**	**Arthrides**
Without limited opening 52% With limited opening 15%	With reduction 41% Without reduction 7%	Arthralgia 36% Osteoarthritis 5% Osteoarthrosis 3%

Percentage of patients with	One Diagnosis	Two Diagnoses	Three Diagnoses
	52%	36%	12%

As regards treatment, the intention of this study was not to influence the treatment plan; this was left entirely up to the treating consultant. Treatments included counselling (in every case); self-administered physiotherapy (in most cases); medications in the form of painkillers, anti-inflammatory agents or anti-depressants; and occlusal treatment by means of soft splint, hard splint or occlusal adjustment. In most cases a combination of two or more treatments was prescribed. When necessary, patients were also referred to a physiotherapist, a pain specialist, or a psychiatrist, though this was uncommon.

All 110 patients completed the data forms prior to any intervention or advice from clinical staff. In addition, ten of them also received these by post one week before their initial appointment. This was necessary for the stability (test-retest) analysis.

One hundred patients were followed-up. Of these, 66 also completed the review at least three months after their initial appointment. The review was performed at patients' follow-up appointments or by post. These data are not presented in this thesis but are part of future work.

6.4.2.2. The control group

A group of TMD-free individuals was needed to test the discriminant validity of the intended short version (OHIP/TMD-I). This group was age (10 year band) and sex matched to patients (Figure 6.2). It consisted of people with no history of TMD who were either attending the Dental Hospital to accompany patients or who were non-clinical staff or non-dental students at the University of Newcastle. Inclusion was based on pre-determined screening questions (Appendix 10) to exclude a history of TMD or a perceived need for treatment for pain or problems in the masticatory muscles or TMJs. The controls completed only the modified version of OHIP.

As recruitment of cases and matched controls was simultaneous, the controls were over-sampled by 16 cases from different age and sex groups. Redundant records were excluded based on the date of recruitment.

6.4.2.3. Estimating the sample size

The sample size was decided initially to allow a sufficient difference in OHIP mean score to be detected between TMD patients and the normal population, and also between patients' scores before and after treatment. Since patients after treatment are anticipated to have residual symptoms that lead to them scoring more on OHIP than the normal population, the difference between them and patients prior treatment is expected to be smaller than that between TMD patients and controls. This, consequently, requires a bigger sample to capture the smaller difference. Therefore, the bigger sample size was used, i.e. which allows the detection of a score difference before and after treatment.

This analysis was based on Altman's nomogram (Altman, 1991; Petrie *et al.*, 2002) and performed using the statistical programme Minitab[1]. To perform this analysis the expected true difference before and after treatment in OHIP mean score, as well as the standard deviation of the TMD patients pre-treatment had to be determined. Since there was no study in the literature that had established these figures in an English speaking population, data from the German population (John *et al.*, 2002) was used as the best estimate. These values were overall OHIP score = 16.7 (SD = 30.0).

In Minitab the formula for comparing paired groups was used. At a significance level of 0.05, and a study power of 90% to be achieved, the number of subjects needed to show a difference similar to that reposted in the German OHIP was 89 subjects. However, I had to "over-sample" in order to compensate for the expected sample attrition by patients' dropouts from either the treatment or the study. I therefore recruited 110 patients and a similar number of controls. This increase also allowed for almost all of the RDC groups to be represented in reasonable numbers.

---

[1] MINITAB® Release 13.1. Minitab Inc., Quality Plaza, 1829 Pine Hall Rd, State College PA 16801-3008, USA. URL: www.minitab.com

To ensure the accuracy of this sample size, I calculated this in a different way based on the estimated standard error of the sample. This process specifies 99 cases in each group assuming a 50 % prevalence at a level which would allow a precision of percentage estimate for each item within 10 percentage points (Riffenburgh, 2006). This figure was close to our previous estimation.

6.4.3. Preparatory stage and pilot work

Approvals for this study were sought from both Newcastle NHS Trust and the Local Research Ethics Committee LREC (Newcastle and North Tyneside LREC-1). The study was registered in LREC under reference number 04/Q0905/24. Written informed consent was obtained from all patients and controls. The information sheet and the consent form are presented elsewhere (Appendix 9).

The study was piloted by administering the full package of instruments to 10 patients and 10 controls. Participants were interviewed after completing the questionnaires to evaluate the feasibility and practicality of procedures. No changes were deemed to be necessary at this stage; therefore, pilot data were used in the final analysis.

6.4.4. Data entry and cleaning

Anonymised data were entered onto SPSS 12[1] on a regular basis. All original files were kept in a safe box and separate from registration forms. After the completion of data collection and entry, the entire data set was cross-checked for typos by reading back the code sequence. Only one patient entry was found to be wrong and this was corrected.

[1] SPSS 12.0.1 for Windows, Release 11 Nov 2003. Chicago, Illinois. http://www.spss.com .

Missing data

Missing data for OHIP were unanswered questions. As OHIP does not allow for a separate not-applicable option, items that were not applicable were marked by patients as 0 (never experienced), and therefore were not considered as missing data. For MPI, missing data included unanswered questions or those answered as not applicable.

The study protocol dictated that patients and controls alike were excluded if they had missing data of ≥ 10% of OHIP items (five items). These were excluded during the recruitment stage, and were not counted as part of the final data set. OHIP was the test instrument, and provided that sufficient OHIP data were present substantial analysis was still possible. For other parts of the data form, patients were not excluded during the recruitment stage for missing scores. This is because many had a non painful condition (e.g. disc displacement without reduction); hence a considerable number of items was marked as not-applicable. At the analysis stage, however, MPI items with ≥ 10% missing data were excluded.

For both OHIP and MPI, two methods were considered to replace missing values (below the exclusion threshold of 10%):

1. The score was replaced with the question's mean score across participants, or

2. With a mean for the participant within the same subcategory.

The first method was adopted and data were replaced.

The remaining self-report questions were single-items. Missing data for these included any unanswered question, which were only a few cases. None of these questions or cases with missing data from these questions was excluded. Also replacement of missing data was not possible.

6.4.5. Data analysis: comparing TMD patient and control groups

The significance of the difference for each item score between the two groups was measured. In addition, summary sample measures for the 49 individual item scores were calculated in four ways: *Mean, Median, Prevalence* and *Item-Impact* as described below. Raw differences between the two groups were examined.

6.4.5.1. Significance of difference between patients and controls for item and domain scores

Before testing the significance of differences between the groups, it was first essential to decide whether parametric or non-parametric tests were to be used. The distribution of data was explored for normality and skewness. Kolmogorov-Smirnov test showed that all items' data from both patients and controls were not normally distributed.

Non-parametric tests were the appropriate tests of significance of difference here. This is because data were not normally distributed and because it is recommended to employ non-parametric tests for ordinal data such as the present (Bowling, 2002).

Data for each of the individual items and for the combined scores within each domain were then compared using a non-parametric test (Mann-Whitney U test).

6.4.5.2. Mean item scores

Although the *Mean* is not generally a suitable test for data at the ordinal level (Bowling, 2002), there is an argument for using it as a method of exploring the data and to compare between different groups (Slade, 2002). *Mean*s were calculated for all items for each group using SPSS. Differences between the two groups in *Mean* were measured for each item, and items were ranked in a descending order according to their differences.

6.4.5.3. Median item scores

Median is the most appropriate statistic to describe the central tendency of ordinal data (Bowling, 2002). The *Median* was calculated for each item in each group using SPSS. Differences in *Median* were measured for each item, and items were ranked in a descending order according to their differences.

6.4.5.4. Prevalence of items

Prevalence of a particular impact (i.e. item) is often used in surveys. To measure this, data should be in a dichotomous order, i.e. having or not having the impact. Data in other forms such as ordinal or numerical must be transformed to binary form.

In OHIP, the percentage of participants reporting an impact over a specific threshold has been widely used before (Allen and Locker, 2002; John *et al.*, 2003). Item occurrence was considered as an impact in different ways. In this study, I calculated percentage of subjects in each group reporting "FOVO" (the prevalence of item response fairly or very often) for each item. Differences in %FOVO between the two groups were then measured for each item, and items were ranked in a descending order by their differences.

The cut-off point used in the present study, FOVO, was somewhat stringent in comparisons to previous studies, which used different approaches, e.g. "OFOVO" (occasionally, fairly or very often) or any impact (from hardly ever to very often) (Allen and Locker, 2002). FOVO was favoured as it captures an impact at a level more likely to require treatment.

6.4.5.5. Item-Impact scores for individual items

The *Item-Impact* analysis identifies items that are the most important to the study subjects (Locker and Allen, 2002). In the original method (Juniper *et al.*, 1997), the *Item-Impact* score is calculated for each item by multiplying the proportion of responders experiencing this item

by its mean importance rating. A modification of this method was introduced by Allen and Locker (2002) whereby the *Item-Impact* is obtained by replacing importance by severity. These authors suggested using "*Mean*" as a measure of severity and "% FOVO" as a measure of frequency. Further details about this test, its use in the literature and potential advantages and disadvantages are considered elsewhere (sections 2.4.3.5 and 2.6.6).

I used the modified method to calculate *Item-Impact* for each item in the two groups. I also measured the difference in *Item-Impact* value between the groups and ranked items in a descending order by their differences.

6.4.5.6. Combined domain and full OHIP scores

Domain scores were calculated by summing the item scores that comprised each domain for each participant (cases and controls). The *Mean* score for each domain was then calculated and compared between cases and controls. Since the domains have different numbers of items I calculated and compared the "*Mean* per item" obtained by dividing the domain *Mean* by the number of items. The full OHIP score was also measured by summing all 49 item scores. A *Mean* for the total OHIP score was also calculated for all subjects in both groups.

In addition, the percentage of items reported fairly or very often (FOVO) within each domain was calculated. The results were compared first between domains to find out the domains with largest impact on quality of life, and second between patients and controls.

6.4.6. Identifying candidate items (OHIP/TMD-I)

One of the primary aims of this research was to start the process of generating a shorter and condition-specific version of OHIP for use as an outcome measure, for example in clinical trials. The recommended method to shorten long instruments to produce a condition-specific version is the *Item-Impact* (Locker and Allen, 2002). I used this method to initially identify candidate items (i.e. those scoring highest in the *Item-Impact* analysis). The *Item-Impact*

analysis, however, does not yield measures of clinical meaning, and the cut-off points between TMD-related and unrelated items could not be drawn based on this analysis alone. For this reason, further item-selection criteria were deemed necessary. These were sequential "filters" that each item should pass to be selected, such that selected items should:

1. Show statistically significant differences between the TMD patients and controls
2. Show "clinically meaningful differences" in the Mean and Median between these two groups
3. Contain at least two items from each OHIP domain.

The concept of a "clinically meaningful difference" in scores was explored by using different thresholds for clinical significance in the *Mean* and *Median*. Items which reached the proposed thresholds were identified. Items were selected based on their distribution within the OHIP theoretical framework, and the set of identified items was called OHIP/TMD-I. These were considered as candidates for the final TMD-QOL product.

In addition to the previous criteria, it is essential for the outcome measure that each item shows a degree of sensitivity to change that is clinically meaningful. Testing responsiveness to changes with treatment is not considered in this thesis, but will be part of future analysis.

6.4.7. Exploring validity and reliability of OHIP and OHIP/TMD-I

As described in the literature review (section 1.1.1), the psychometric properties of a new instrument should be established, even if it is derived from a previously validated instrument. It is highly recommended, however, that these properties are tested in a sample other than that used in the derivation phase. The aim of this part of the project was to explore aspects of validity and reliability of the full OHIP and compare these to the psychometric properties of the proposed set of OHIP items (OHIP/TMD-I). It was not the intention to report the establishment of these properties for the new instrument. Further full validation is a substantial piece of work and will be completed at a later stage.

6.4.7.1. Content validity

This type of validity explores whether content of the instrument covers everything it should, and does not include irrelevant elements. This is assessed by members of the target population.

The main thrust of the present work was to shorten the original OHIP, and only maintain items with the most impact. Future qualitative work would help to ensure the new instrument is not missing important themes for TMD patients. Should this be the case, new items will be added and appropriately tested.

In addition, in order to improve this aspect of validity I have slightly modified the questions by adding the word *"jaws"* as a source of the problem listed in the original OHIP as relating to *"teeth, mouth, or dentures"*. This ensured that participants were able to relate the questions to their TMD. I also altered the recall (reference) period to one month, which has proven to be most suitable for TMD outcome assessment (*Study One* and *Study Two* in this project).

Furthermore, I have maintained the hierarchal structure of OHIP (i.e. Locker's model) in order to maximise the content validity of the derived instrument, in agreement with Slade (1997b) and Locker and Allen (2002).

6.4.7.2. Criterion validity

Criterion validity examines the inter-correlation between the investigated instrument and a gold standard measurement of the same concept applied concurrently on the same group. When testing the validity of a new shortened instrument, the original version is considered the gold standard against which the short form can be tested (Slade, 2002).

Correlation between and OHIP/TMD-I and full OHIP was measured by means of Pearson's test.

6.4.7.3. Discriminant validity

The instrument should be able to show statistical and clinically meaningful differences where these are expected:

1. Between TMD patients and controls

2. Before and after treatment.

If the instrument is not able to capture a difference between TMD sufferers and TMD-free subjects, then it is unlikely to capture a difference of a smaller attitude between patients' scores pre- and post-treatment.

The ability of OHIP items to show statistically and clinically significant difference between patients and controls was examined. Statistically significant difference was tested (see 6.4.5.1 and 6.5.2.1). As part of the shortening procedure, items showing no statistical difference between the two groups were not included in OHIP/TMD-I.

Ensuring that all selected items have the ability to show clinically meaningful difference was also part of the distilling process. However, for further examination of this issue, I compared the present results with another study that compared the full OHIP scores between patients and the normal population in Germany (John *et al.*, 2007).

6.4.7.4. Convergent validity

To explore the convergent validity of OHIP and OHIP/TMD-I in the TMD population, these instruments were checked against another measure of chronic pain's impact on quality of life - the Multidimensional Pain Inventory (MPI), and two other self-report measures of pain (visual analogue scales, VAS1 and VAS2); and against a global rating of oral health.

Only section one of the MPI was used in this study, which consists of 20 items. Four of these items showed a considerable amount of missing data (see page 199) and had to be excluded from further analyses. In the following sections, the term MPI refers to the 16

remaining items. As described in greater detail in the results section, it was necessary to reverse the positive scale for three items where a high score indicates improved QOL.

Summary scores for MPI and the other above measures were calculated for the patient group including *Mean*, Standard error of mean and *Median*.

The convergent validity was expressed as Pearson's correlation coefficients for OHIP49 and OHIP/TMD-I cross tabulated against the other measures described above.

6.4.7.5. Test-retest reliability

I examined the stability of OHIP and its proposed short form (OHIP/TMD-I), i.e. its ability to reproduce a TMD patient's scores on two occasions given the condition was "relatively" stable. Participants in this stage of the study completed OHIP twice, one week before their first appointments, and on the day of first appointment.

Twenty patients were initially sent an invitation letter to take part in the study which included the information sheet, and all the self-report single-items and questionnaires that other patients received in their first appointment, including OHIP, MPI, and pain VASs.

Ten patients agreed to take part. Their scores of full OHIP and OHIP/TMD-I were each compared between the two occasions using the Intra-class Correlation Coefficient, and using the Wilcoxon test to explore any significant difference between them.

6.4.7.6. Internal consistency reliability

It is essential to measure the internal consistency for existing instruments, e.g. scales, with established internal consistency when these are used in a population other than that for which they were designed (Litwin, 1995). In this study I measured Cronbach' α coefficient for two scales: the full OHIP and the proposed short form.

Internal consistency reliability or the "Homogeneity" test explores the harmony between the instrument's elements and investigates whether the questionnaires' individual items are

measuring the same underlying construct. *Cronbach' α* coefficient is the most commonly used indicator of homogeneity, which has to reach a value of 0.7 for the questionnaire to have acceptable internal consistency (Nunnally and Berbstein, 1994). To improve the internal consistency of a scale, it is recommended to exclude redundant items, i.e. those which are not measuring the same construct and are reducing the homogeneity of the scale. These are excluded based on their value of Corrected Item-Total Correlation obtained as part of the Cronbach' α analysis. Items with a value of less than 0.3 are considered redundant (Pallant, 2005).

6.5. Results

6.5.1. Missing data

Of those who met the inclusion criteria, only three controls and ten patients had missing data in the OHIP forms. This ranged from one to three items for each subject, except one case with four missing items.

Missing items varied, but the ones with the highest frequency were items nine and eighteen, with a frequency of missing answers of five and three respectively. Both questions are related to denture fitting and comfort.

For MPI, the following items were reported with more than 10% missing data, and were therefore excluded from further analyses:

MPI-05: How supportive is the spouse? Missing in 13 patients

MPI-10: How worried is the spouse about you? Missing in12 patients

MPI-14: Pain changed the amount of satisfaction from work. Missing in 21 patients

MPI-15: How attentive is your spouse to your pain problem? Missing in12 patients.

There were a further three items with missing data. These items were missing from one to six patients, and were replaced by means of patients' scores. In addition, one patient (No. 31) answered only five MPI questions. She was therefore excluded from the comparison between OHIP and MPI.

Data from other self-report measures showed very few missing values. Two patients did not complete both VAS1 and VAS2; one of whom did not also give global rating of oral health.

6.5.2. Differences between TMD patient and control groups

6.5.2.1. Significance of difference

All item scores were higher for patients than controls. Forty one of the 49 items showed significant differences between groups at $p<0.01$ [Mann-Whitney test (Table 6.2)]. The eight items that were not significantly different were not obviously related to TMD (e.g. denture problems, aesthetics).

OHIP domain scores were calculated by summing scores from all items in each domain – see section 6.5.2.6). All seven OHIP domain scores showed very highly significant differences in scores between patients and controls ($p<0.001$).

Table 6.2. Mann-Whitney test and significance of score differences between patients and controls
Items are ranked in their original order in OHIP

Item No.	OHIP item	Mann-Whitney	Asymp. Sig. (2-tailed)
01	Difficulty chewing foods	836.5	0.00000
02	Trouble pronouncing words	3546.5	0.00000
03	Tooth doesn't look right	5872.5	0.67100*
04	Appearance affected	4822.5	0.00400
05	Breath has been stale	4901.5	0.00900
06	Sense of taste has worsened	5041	0.00700
07	Food catching in teeth	5398.5	0.22300*
08	Digestion has worsened	4670	0.00000
09	Dentures have not been fitting	5752	0.90600*
10	Painful aching in mouth	1488	0.00000
11	Sore jaw	430.5	0.00000
12	Headaches	1985	0.00000
13	Sensitive teeth	4734.5	0.00500
14	Toothache	4407	0.00000
15	Painful gums	4220.5	0.00000
16	Uncomfortable to eat foods	1448	0.00000
17	Sore spots in mouth	4684	0.00200
18	Uncomfortable dentures	5763	0.55700*
19	Worried by jaw/dental problems	1077.5	0.00000
20	Have been self conscious	3706	0.00000
21	Dental problems made you miserable	1955	0.00000
22	Uncomfortable about the appearance	4661.5	0.00100
23	Felt tense	1639	0.00000
24	Speech has been unclear	4392.5	0.00000
25	People misunderstood some words	4827	0.00000
26	Less flavour in food	5259	0.01700*
27	Unable to brush your teeth properly	3476.5	0.00000
28	Avoid eating some foods	1560	0.00000
29	Diet has been unsatisfactory	3916	0.00000
30	Unable to eat with dentures	5807	0.52700*
31	Avoided smiling	5116	0.01200*
32	Have to interrupt meals	2629	0.00000
33	Sleep has been interrupted	2534.5	0.00000
34	Has been upset	2166	0.00000
35	Found it difficult to relax	1996	0.00000
36	Felt depressed	3113	0.00000
37	Concentration has been affected	2313.5	0.00000
38	Has been a bit embarrassed	4156	0.00000
39	Avoided going out	5080	0.00100
40	Less tolerant of spouse or family	3612	0.00000
41	Trouble getting on with other people	4712	0.00000
42	Irritable with other people	3309.5	0.00000
43	Difficulty doing your usual jobs	4240.5	0.00000
44	General health has worsened	3765.5	0.00000
45	Suffered any financial loss	5772.5	0.22700*
46	Unable to enjoy others company	4355	0.00000
47	Life in general was less satisfying	2971.5	0.00000
48	Totally unable to function	5052	0.00000
49	Unable to work to full capacity	4200	0.00000

*** Highlighted in red are items showing no statistically significant difference**

6.5.2.2. Mean

As shown in Table 6.3 the highest *Mean* score for patients was 3.05 (SE= 1.00) for item 11 (sore jaw), while the lowest was 0.19 (0.06) for item 45 (suffered financial loss). For controls, these figures were much lower, with the highest *Mean* score of just 1.59 (0.10) for item 7 (food catching in teeth), and lowest *Mean* score of 0.05 (0.03) for item 48 (totally unable to function).

Differences in *Mean* scores between patients and controls ranged between 0.03 and 2.7 units, with an average difference of 0.89 unit per item (Table 6.3 and Figure 6.4 A).

6.5.2.3. Median

Out of a possible five units (0-4), *Median* scores ranged between 0 and 3 in the patient group. Table 6.4 shows the range of *Median* scores for all 49 items. For patients, there were three items with a *Median* of 3, and a further 13 items with a *Median* of 2. The remaining items scored either 1 or 0. These figures were very different from the controls' scores where only one item (No. 7: food catching in teeth) showed a *Median* of 2, and another (No. 13: sensitive teeth) a *Median* of 1. The *Median* for the remaining items was 0 reflecting no impact on quality of life of the controls during the last month.

Differences in *Median* values between patients and controls varied between 0 and 3 (Table 6.4 and Figure 6.4 B). Since *Median* values for the controls were 0 for almost all items, the difference in *Median* values between patients and controls reflected mainly the patients' figures. As mentioned above, the two exceptions were items 7 and 13.

Table 6.3. Differences in item *Mean* scores between patients and controls

Items are ranked in a descending order according to the difference

Item No.	OHIP item	Patients		Controls		Difference in *Mean**
		Mean	SE	*Mean*	SE	
11	Sore jaw	3.05	0.1	0.35	0.063	2.70
01	Difficulty chewing foods	2.61	0.102	0.49	0.073	2.12
19	Worried by jaw/dental problems	2.53	0.107	0.45	0.075	2.08
28	Avoid eating some foods	2.29	0.13	0.36	0.073	1.93
10	Painful aching in mouth	2.55	0.116	0.65	0.08	1.90
16	Uncomfortable to eat foods	2.45	0.112	0.58	0.081	1.87
12	Headaches	2.09	0.139	0.31	0.068	1.78
23	Felt tense	1.96	0.123	0.24	0.055	1.72
21	Dental problems made you miserable	2.04	0.129	0.39	0.067	1.65
35	Found it difficult to relax	1.83	0.128	0.26	0.062	1.57
33	Sleep has been interrupted	1.76	0.146	0.24	0.059	1.52
34	Has been upset	1.72	0.128	0.26	0.064	1.46
37	Concentration has been affected	1.55	0.123	0.18	0.052	1.37
32	Have to interrupt meals	1.44	0.127	0.15	0.044	1.29
36	Felt depressed	1.27	0.128	0.17	0.051	1.10
20	Have been self conscious	1.65	0.141	0.65	0.095	1.00
27	Unable to brush your teeth properly	1.33	0.127	0.34	0.064	0.99
42	Irritable with other people	1.15	0.118	0.16	0.051	0.99
47	Life in general was less satisfying	1.04	0.11	0.1	0.045	0.94
40	Less tolerant of spouse or family	1.07	0.122	0.17	0.055	0.90
44	General health has worsened	0.87	0.111	0.13	0.05	0.74
29	Diet has been unsatisfactory	0.83	0.11	0.1	0.034	0.73
02	Trouble pronouncing words	0.85	0.102	0.14	0.051	0.71
15	Painful gums	1.33	0.122	0.64	0.082	0.69
49	Unable to work to full capacity	0.78	0.11	0.1	0.036	0.68
38	Has been a bit embarrassed	1	0.115	0.34	0.07	0.66
43	Difficulty doing your usual jobs	0.75	0.109	0.14	0.051	0.61
46	Unable to enjoy others company	0.74	0.105	0.14	0.044	0.60
17	Sore spots in mouth	1.26	0.128	0.68	0.084	0.58
14	Toothache	1.36	0.115	0.8	0.09	0.56
22	Uncomfortable about the appearance	1.05	0.126	0.49	0.083	0.56
24	Speech has been unclear	0.66	0.102	0.12	0.048	0.54
04	Appearance affected	1.19	0.133	0.69	0.106	0.50
13	Sensitive teeth	1.71	0.122	1.24	0.114	0.47
31	Avoided smiling	0.8	0.12	0.36	0.072	0.44
05	Breath has been stale	1.08	0.109	0.66	0.079	0.42
08	Digestion has worsened	0.56	0.086	0.16	0.044	0.40
39	Avoided going out	0.49	0.093	0.1	0.032	0.39
41	Trouble getting on with other people	0.46	0.084	0.07	0.04	0.39
25	People misunderstood some words	0.51	0.087	0.13	0.037	0.38
06	Sense of taste has worsened	0.65	0.097	0.28	0.056	0.37
48	Totally unable to function	0.39	0.09	0.05	0.027	0.34
26	Less flavour in food	0.45	0.08	0.18	0.045	0.27
07	Food catching in teeth	1.78	0.123	1.59	0.098	0.19
30	Unable to eat with dentures	0.2	0.073	0.08	0.032	0.12
18	Uncomfortable dentures	0.2	0.072	0.11	0.043	0.09
45	Suffered any financial loss	0.19	0.058	0.1	0.039	0.09
09	Dentures have not been fitting	0.22	0.076	0.14	0.051	0.08
03	Tooth doesn't look right	0.82	0.108	0.79	0.113	0.03

*** Highlighted in green and orange are items with difference in *Mean* of ≥2 units, ≥ 1 unit respectively**

Table 6.4. Differences in item *Median* scores between patients and controls

Items are ranked in a descending order according to the difference in *Median*

Item No.	OHIP item	Patients	Controls	Difference in *Median**
10	Painful aching in mouth	3	0	3
11	Sore jaw	3	0	3
19	Worried by jaw/dental problems	3	0	3
01	Difficulty chewing foods	2	0	2
12	Headaches	2	0	2
16	Uncomfortable to eat foods	2	0	2
20	Have been self conscious	2	0	2
21	Dental problems made you miserable	2	0	2
23	Felt tense	2	0	2
28	Avoid eating some foods	2	0	2
33	Sleep has been interrupted	2	0	2
34	Has been upset	2	0	2
35	Found it difficult to relax	2	0	2
37	Concentration has been affected	2	0	2
14	Toothache	1	0	1
15	Painful gums	1	0	1
27	Unable to brush your teeth properly	1	0	1
32	Have to interrupt meals	1	0	1
36	Felt depressed	1	0	1
42	Irritable with other people	1	0	1
47	Life in general was less satisfying	1	0	1
17	Sore spots in mouth	1	0	1
04	Appearance affected	1	0	1
13	Sensitive teeth	2	1	1
05	Breath has been stale	1	0	1
38	Has been a bit embarrassed	0.5	0	0.5
02	Trouble pronouncing words	0	0	0
08	Digestion has worsened	0	0	0
24	Speech has been unclear	0	0	0
25	People misunderstood some words	0	0	0
29	Diet has been unsatisfactory	0	0	0
40	Less tolerant of spouse or family	0	0	0
41	Trouble getting on with other people	0	0	0
43	Difficulty doing your usual jobs	0	0	0
44	General health has worsened	0	0	0
46	Unable to enjoy others company	0	0	0
48	Totally unable to function	0	0	0
49	Unable to work to full capacity	0	0	0
22	Uncomfortable about the appearance	0	0	0
39	Avoided going out	0	0	0
06	Sense of taste has worsened	0	0	0
31	Avoided smiling	0	0	0
26	Less flavour in food	0	0	0
07	Food catching in teeth	2	2	0
45	Suffered any financial loss	0	0	0
30	Unable to eat with dentures	0	0	0
18	Uncomfortable dentures	0	0	0
03	Tooth doesn't look right	0	0	0
09	Dentures have not been fitting	0	0	0

*** Highlighted in green, orange and yellow are items with difference in *Median* of 3, 2 and 1units respectively**

6.5.2.4. Prevalence

The prevalence of individual items varied widely (Table 6.5). Prevalence reported here is based on the percentage of subjects reporting impacts occurring fairly often or very often (% FOVO), in other words, on a high level of frequency. The most prevalent item amongst TMD patients was item 11 (sore jaw) with a % FOVO of 67%. At the other end of the scale, item 45 (suffered financial loss) showed the lowest frequency with a prevalence of 1 %.

For controls, the most prevalent item was item 7 (food catching in teeth) with a % FOVO OF just 13%, while thirteen items had a % FOVO OF zero. Examples of these are: sore jaw, problems with digestion or with dentures, etc.

Differences in %FOVO between patients and controls ranged widely across items between 1% and 67% as presented in Table 6.5 and Figure 6.4 C.

Table 6.5. Differences in item *Prevalence* (%FOVO) between patients and controls

Items are ranked in a descending order according to the difference patient: control

Item No.	OHIP item	Patients	Controls	Difference in % FOVO
11	Sore jaw	67	0	67
10	Painful aching in mouth	53	1	52
19	Worried by jaw/dental problems	53	2	51
01	Difficulty chewing foods	46	1	45
16	Uncomfortable to eat foods	47	4	44
12	Headaches	42	1	41
28	Avoid eating some foods	41	2	39
21	Dental problems made you miserable	37	1	36
23	Felt tense	34	1	33
33	Sleep has been interrupted	34	1	33
35	Found it difficult to relax	30	1	29
34	Has been upset	26	3	24
32	Have to interrupt meals	25	1	24
20	Have been self conscious	28	5	23
37	Concentration has been affected	22	1	21
27	Unable to brush your teeth properly	19	0	19
15	Painful gums	21	3	18
07	Food catching in teeth	30	13	17
17	Sore spots in mouth	18	2	16
36	Felt depressed	16	1	15
13	Sensitive teeth	25	12	13
05	Breath has been stale	13	0	13
31	Avoided smiling	14	2	12
40	Less tolerant of spouse or family	13	1	12
42	Irritable with other people	13	1	12
43	Difficulty doing your usual jobs	12	1	11
47	Life in general was less satisfying	12	1	11
22	Uncomfortable about the appearance	14	3	11
38	Has been a bit embarrassed	13	2	11
14	Toothache	14	4	10
49	Unable to work to full capacity	10	0	10
44	General health has worsened	10	0	10
04	Appearance affected	17	8	9
29	Diet has been unsatisfactory	9	0	9
46	Unable to enjoy others company	8	1	7
06	Sense of taste has worsened	7	0	7
02	Trouble pronouncing words	9	3	7
48	Totally unable to function	6	0	6
24	Speech has been unclear	7	2	5
39	Avoided going out	5	0	5
08	Digestion has worsened	5	0	5
25	People misunderstood some words	5	0	5
30	Unable to eat with dentures	4	0	4
41	Trouble getting on with other people	5	1	4
09	Dentures have not been fitting	4	1	3
26	Less flavour in food	3	0	3
18	Uncomfortable dentures	3	1	2
03	Tooth doesn't look right	10	9	1
45	Suffered any financial loss	1	0	1

6.5.2.5. Item-Impact

Similar to the previous analyses, the *Item-Impact* (Table 6.6) showed that in comparison to the control group, TMD patients reported a high impact of their condition on their quality of life and a higher degree of differences across items.

Unsurprisingly, the same "extreme" items were identified. For example, item 11 (sore jaw) had the largest *Item-Impact* (205.18) in patients, while item 45 (suffered financial loss) showed the lowest impact (0.17). The controls' figures were again very much lower than the corresponding patients'; the question with highest item impact was again No. 7: food catching in teeth (20.24), and thirteen items had a value of 0 (reflecting zero *Prevalence*).

Differences in *Item-Impact* values between patients and controls ranged widely between 0.2 and 205.2 as shown in Table 6.6.

Figure 6.3. *Item-Impact* scores for individual items for patients and controls

Items are sorted in a descending order of patients' scores

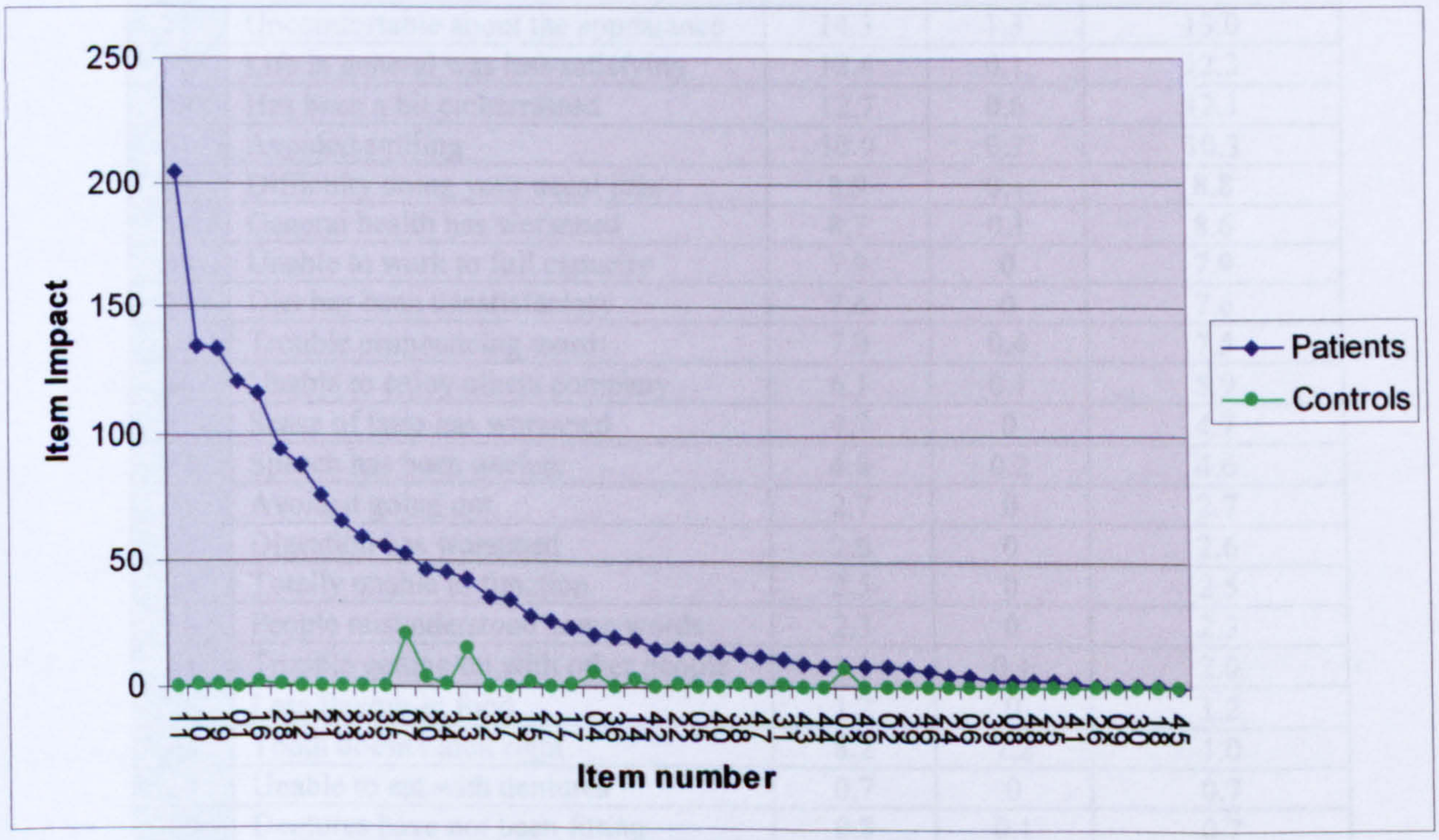


Table 6.6. Differences in *Item-Impact* values between patients and controls

Items are ranked in a descending order according to the difference

Item No.	OHIP item	Patients	Controls	Difference in *Item-Impact*
11	Sore jaw	205.2	0	205.2
10	Painful aching in mouth	134.5	0.6	133.9
19	Worried by jaw/dental problems	133.4	0.8	132.6
01	Difficulty chewing foods	120.8	0.5	120.4
16	Uncomfortable to eat foods	115.8	2.1	113.7
28	Avoid eating some foods	93.7	0.7	93.0
12	Headaches	87.4	0.3	87.1
21	Dental problems made you miserable	76.0	0.4	75.7
23	Felt tense	65.9	0.2	65.7
33	Sleep has been interrupted	59.2	0.2	59.0
35	Found it difficult to relax	54.9	0.2	54.7
34	Has been upset	45.4	0.7	44.6
20	Have been self conscious	46.5	3.6	43.0
32	Have to interrupt meals	35.4	0.1	35.2
37	Concentration has been affected	33.8	0.2	33.7
07	Food catching in teeth	52.7	20.2	32.5
13	Sensitive teeth	42.4	14.7	27.7
15	Painful gums	27.8	1.8	26.1
27	Unable to brush your teeth properly	25.4	0	25.4
17	Sore spots in mouth	22.9	1.2	21.7
36	Felt depressed	19.6	0.2	19.5
14	Toothache	18.7	2.9	15.8
04	Appearance affected	20.6	5.7	14.9
42	Irritable with other people	14.6	0.2	14.5
05	Breath has been stale	13.8	0	13.8
40	Less tolerant of spouse or family	13.6	0.2	13.5
22	Uncomfortable about the appearance	14.3	1.3	13.0
47	Life in general was less satisfying	12.4	0.1	12.3
38	Has been a bit embarrassed	12.7	0.6	12.1
31	Avoided smiling	10.9	0.7	10.3
43	Difficulty doing your usual jobs	8.9	0.1	8.8
44	General health has worsened	8.7	0.1	8.6
49	Unable to work to full capacity	7.9	0	7.9
29	Diet has been unsatisfactory	7.6	0	7.6
02	Trouble pronouncing words	7.9	0.4	7.5
46	Unable to enjoy others company	6.1	0.1	5.9
06	Sense of taste has worsened	4.7	0	4.7
24	Speech has been unclear	4.8	0.2	4.6
39	Avoided going out	2.7	0	2.7
08	Digestion has worsened	2.6	0	2.6
48	Totally unable to function	2.5	0	2.5
25	People misunderstood some words	2.3	0	2.3
41	Trouble getting on with other people	2.1	0.1	2.0
26	Less flavour in food	1.2	0	1.2
03	Tooth doesn't look right	8.2	7.2	1.0
30	Unable to eat with dentures	0.7	0	0.7
09	Dentures have not been fitting	0.8	0.1	0.7
18	Uncomfortable dentures	0.6	0.1	0.5
45	Suffered any financial loss	0.2	0	0.2

6.5.2.6. Combined domain and full OHIP scores

On a possible range of 0-245 points on the OHIP total score, patients scored 60.54 on average, while the controls' average was only 17.05. The average item score in general measured 1.24 for patients and 0.35 for controls.

As shown in Table 6.7, OHIP domains exhibited different scores both within the same case group (patients or controls) and between the two groups. Reported here is the *Mean per item*, measured by dividing the mean score for each domain by number of items in the domain.

Table 6.7. Average scores for OHIP domains and total OHIP in patients and controls

Domains are sorted according to difference between the two groups.

	No of items	*Mean* per item		Difference in *Mean per item (substitution)*
		Patients	Controls	
Psychological discomfort	5	1.8	0.4	1.4
Psychological disability	6	1.5	0.2	1.3
Physical pain	9	1.8	0.6	1.2
Physical disability	9	0.9	0.2	0.7
Social disability	5	0.8	0.1	0.7
Handicap	6	0.7	0.1	0.6
Function	9	1	0.5	0.5
Overall OHIP score	49	1.3	0.4	0.9

For TMD cases, the *Mean per item* ranged between 0.7 for handicap and 1.8 for psychological discomfort and for physical pain. For controls, the scores were much lower, with *Mean per item* ranging from 0.1 for social disability and handicap to 0.6 for physical pain. When comparing between TMD sufferers and controls, the largest "*Mean per item*" differences were found for psychological discomfort, psychological disability and physical pain (around 1.3 unit), followed by the other domains with a difference of around 0.6 unit. (Table 6.7).

Patients reporting items with fairly or very often (FOVO) impact on their lives outnumbered controls 10 times overall (Table 6.8). Again OHIP domains showed differences within and between groups. For example, the difference between patients and controls was largest for *physical disability* (20 times), and lowest for *functional limitation* (four times higher).

Table 6.8. Percentage reporting FOVO impact to one or more items across domains and groups
Domains are sorted according to difference between the two groups.

	% FOVO		Difference between the two groups (division)
	Patients	Controls	
Physical disability	14.0%	0.7%	20
Psychological disability	23.3%	1.4%	17
Handicap	7.9%	0.5%	16
Psychological discomfort	33.1%	2.4%	14
Social disability	9.5%	0.7%	14
Physical pain	32.3%	2.9%	11
Functional limitation	15.6%	3.9%	4
Overall OHIP score	20.0%	2.0%	10

6.5.3. Candidate items (OHIP/TMD-I)

The *Item-impact* analysis was not sufficient on its own to determine candidate items. Items scoring the largest difference between patients and controls in the *Item-impact* analysis were selected if they passed three other sequential "filters" set out in the following three sections.

6.5.3.1. Statistical significance of difference

The first inclusion criterion was that items should show a statistically significant difference between TMD patients and controls. Eight items (Table 6.9), previously identified as showing no statistical differences (Table 6.2), were eliminated.

Table 6.9. Items with no statistically significant difference between patients and controls

Items are sorted in their original order in OHIP

Item No.	OHIP item	Mann-Whitney U	Asymp. Sig. (2-tailed)
03	Tooth doesn't look right	5873	0.67
07	Food catching in teeth	5399	0.22
09	Dentures have not been fitting	5752	0.91
18	Uncomfortable dentures	5763	0.56
26	Less flavour in food	5259	0.02
30	Unable to eat with dentures	5807	0.53
31	Avoided smiling	5116	0.01
45	Suffered any financial loss	5773	0.23

6.5.3.2. Clinical thresholds of differences

The second filter to distinguish between TMD related and unrelated items dictated a "clinically meaningful" difference between patients and controls essential for each item. Although different summary scores were measured, clinically meaningful units were only available from the *Mean* and *Median* analyses.

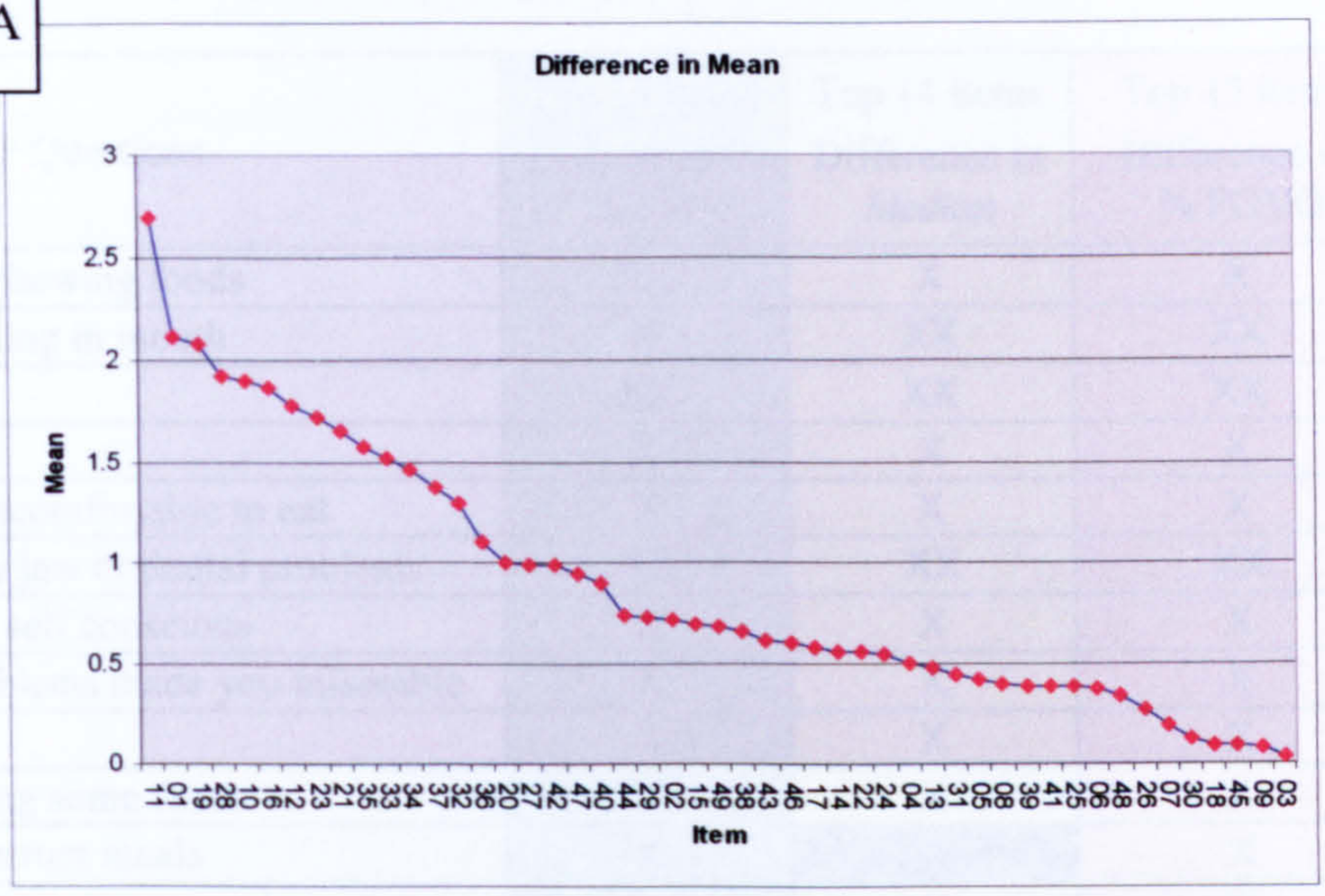
Using the *Mean* (Figure 6.4 A, and Table 6.10), the top 3 ranked items scored a difference between patients and controls of 2 or more units. A further 13 items scored a difference of 1 or more unit. A difference of 1 unit between patients and controls was set as a threshold to select candidate items. This difference is arbitrary, but would probably be clinically meaningful for clinicians, representing for instance the difference between an impact reported as occurring occasionally and fairly often. Clearly, fractions of a unit, i.e. values below 1 unit or between 1 and 2 units would be difficult to interpret. Also, a cut-off point of 2 or more units' difference in the *Mean* results in the identification of only 3 items, which is unlikely to lead to a valid instrument.

For *Median* (Figure 6.4 B and Table 6.10), a difference of 3 units was found for 3 questions with 11 more items showing a difference of 2 units. All these 14 items were amongst the 16 which exceeded the threshold for *Mean*.

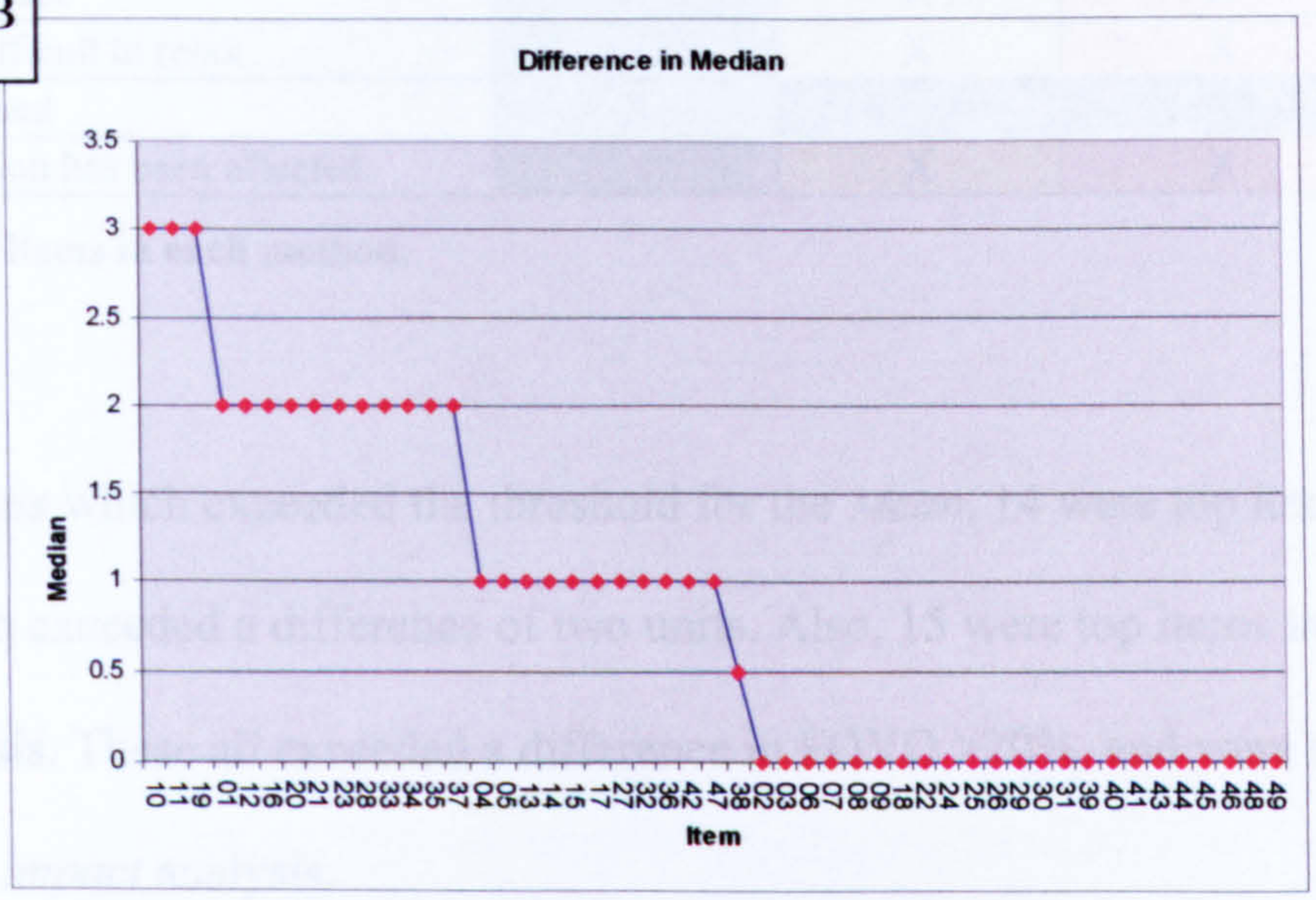
Figure 6.4. Item differences between patients and controls in *Mean*, *Median* and *Prevalence*

Items are sorted in a descending order.

A



B



C

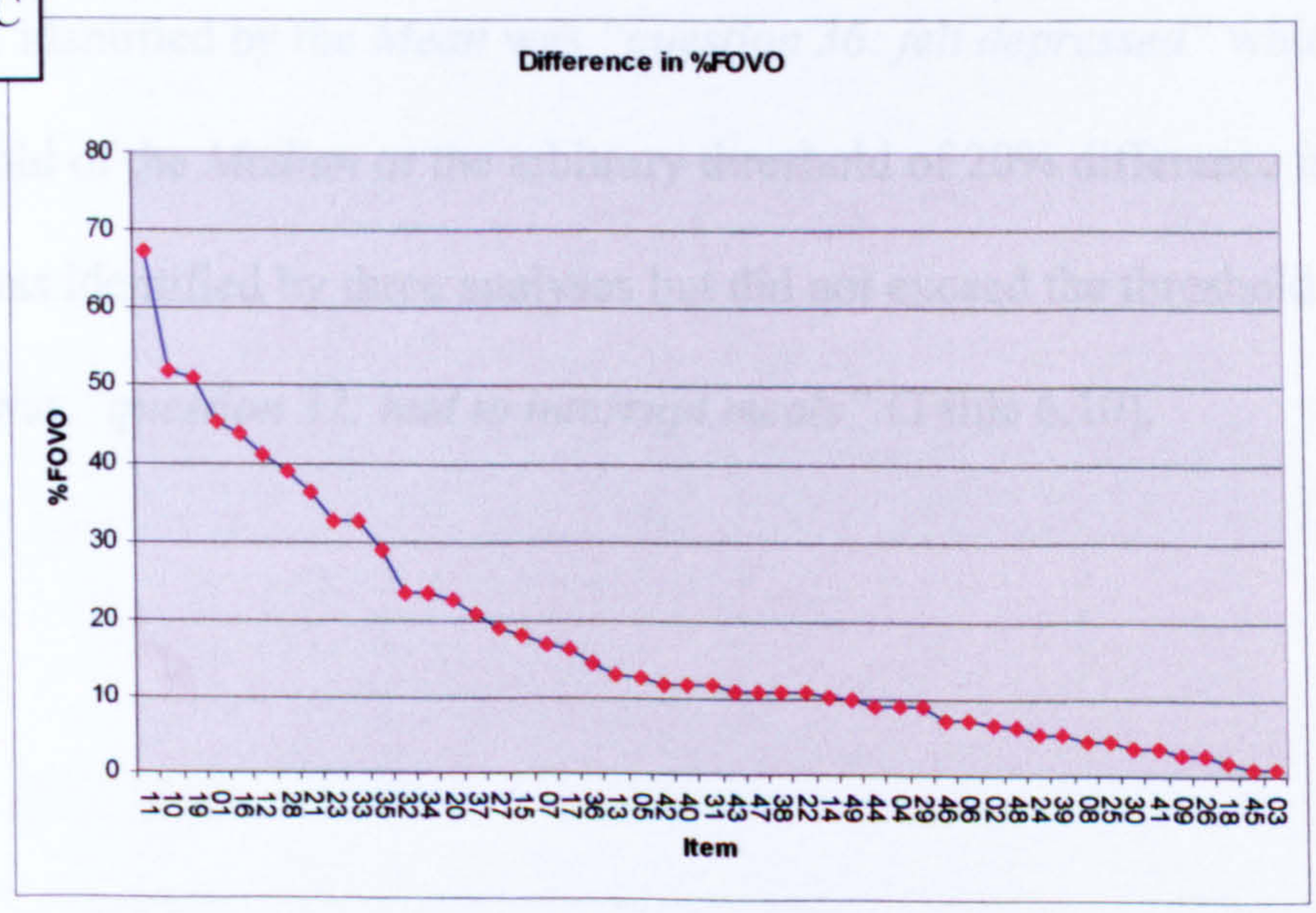


Table 6.10. Items showing the largest difference between patients and controls in *Mean*, *Median*, % FOVO and *Item-Impact*– items are sorted in their OHIP original order

OHIP Questions		Top 16 items Difference in *Mean*	Top 14 items Difference in *Median*	Top 15 items Difference in % FOVO	Top15 items Difference in *Item-Impact*
Q 01	Difficulty chewing foods	XX	X	X	X
Q 10	Painful aching in mouth	X	XX	XX	XX
Q 11	Sore jaw	XX	XX	XX	XX
Q 12	Headaches	X	X	X	X
Q 16	Found it uncomfortable to eat	X	X	X	X
Q 19	Worried by jaw or dental problems	XX	XX	XX	XX
Q 20	Have been self conscious	X	X	X	X
Q 21	Dental problems made you miserable	X	X	X	X
Q 23	Felt tense	X	X	X	X
Q 28	Avoid eating some foods	X	X	X	X
Q 32	Had to interrupt meals	X		X	X
Q 33	Sleep has been interrupted	X	X	X	X
Q 34	Has been upset	X	X	X	X
Q 35	Found it difficult to relax	X	X	X	X
Q 36	Felt depressed	X			
Q 37	Concentration has been affected	X	X	X	X

XX marks top three items in each method.

Of the 16 items which exceeded the threshold for the *Mean*, 14 were top items in the *Median* test which exceeded a difference of two units. Also, 15 were top items in the *Prevalence* analysis. These all exceeded a difference in FOVO >20%, and were the top 15 items in the *Item-Impact* analysis.

The 16th item identified by the *Mean* was *"question 36: felt depressed"* which did not exceed the threshold of the *Median* or the arbitrary threshold of 20% difference in FOVO. The item which was identified by three analyses but did not exceed the threshold for the *Median* analysis was *"question 32: had to interrupt meals"* (Table 6.10).

6.5.3.3. Representation of OHIP domains

This was the third criterion for candidate items. The study protocol dictated that the short form of OHIP should contain at least two items from each OHIP domain. The 16 items identified so far represent only five of OHIP's seven domains, as shown in Table 6.11. The two non-represented domains were *Social disability* and *Handicap*.

Table 6.11. Domains of the items with largest difference between patients and controls

Domain	Item No.	Question
Functional limitation	Q 01	Difficulty chewing foods
Physical pain	Q 10	Painful aching in mouth
	Q 11	Sore jaw
	Q 12	Headaches
	Q 16	Found it uncomfortable to eat
Psychological discomfort	Q 19	Worried by jaw or dental problems
	Q 20	Have been self conscious
	Q 21	Dental problems made you miserable
	Q 23	Felt tense
Physical disability	Q 28	Avoid eating some foods
	Q 32	Had to interrupt meals
Psychological disability	Q 33	Sleep has been interrupted
	Q 34	Has been upset
	Q 35	Found it difficult to relax
	Q 36	Felt depressed
	Q 37	Concentration has been affected

In this stage two items which had the highest difference in *Item-Impact* score between patients and controls from each of the two unrepresented domains were added (see Table 6.12). These four items were also top in their corresponding domains using the *Mean* and *Prevalence* analyses analysis. As for *Median*, only one item in each of these two domains showed a difference of one unit. The remaining items were all equal with a *Median* of zero.

Table 6.12. Additional items and their ranking in the four analyses*

Domain	Item No.	Item	*Item-Impact rank*	*Mean rank*	%FOVO *rank*	*Median rank*
Social disability	Q 40	Less tolerant of spouse or family	1	1	1	N/A**
	Q 42	Irritable with other people	2	2	2	1
Handicap	Q 47	Life in general was less satisfying	1	1	1	1
	Q 44	General health has worsened	2	2	2	N/A**

* **Difference patients: controls**

** **This item has the same *Median* value as the other items in the corresponding domains**

At this stage, all domains were represented by at least two items, except the functional limitation domain with only one item (Table 6.13). The second top item in that domain was No. 7 (food catching in teeth). It did not show, however, a statistically significant difference between patients and controls. The remaining items in this domain showed either no statistically significant different between the case groups or were clearly non-related to TMD and had a very low item-impact. Therefore, no additional items were included from this domain. The final set of 20 items, so called OHIP/TMD-I, is shown in Table 6.14.

Table 6.13. Summary scores and domains for OHIP/TMD-I*

Item No.	Question	*Mean*	*Median*	*Prevalence (%FOVO)*	*Item-Impact*	Domain
Q 01	Difficulty chewing foods	2.12	2	45%	120.4	*Functional limitation*
Q 10	Painful aching in mouth	1.90	3	52%	133.9	*Physical pain*
Q 11	Sore jaw	2.70	3	67%	205.2	
Q 12	Headaches	1.78	2	41%	87.1	
Q 16	Found it uncomfortable to eat	1.87	2	44%	113.7	
Q 19	Worried by jaw or dental problems	2.08	3	51%	132.6	*Psychological discomfort*
Q 20	Have been self conscious	1.00	2	23%	43	
Q 21	Dental problems made you miserable	1.65	2	36%	75.7	
Q 23	Felt tense	1.72	2	33%	65.7	
Q 28	Avoid eating some foods	1.93	2	39%	93	*Physical disability*
Q 32	Had to interrupt meals	1.29	1	24%	35.2	
Q 33	Sleep has been interrupted	1.52	2	33%	59	*Psychological disability*
Q 34	Has been upset	1.46	2	24%	44.6	
Q 35	Found it difficult to relax	1.57	2	29%	54.7	
Q 36	Felt depressed	1.10	1	15%	19.5	
Q 37	Concentration has been affected	1.37	2	21%	33.7	
Q 40	Less tolerant of spouse or family	0.90	0	12%	13.5	*Social disability*
Q 42	Irritable with other people	0.99	1	12%	14.5	
Q 44	General health has worsened	0.74	0	9%	8.6	*Handicap*
Q 47	Life in general was less satisfying	0.94	1	11%	12.3	

* **Difference patients: controls**

Table 6.14. OHIP/TMD-I

Question Number	Question	Domain
Q 01	Have you had difficulty chewing any foods because of problems with your jaws, teeth, mouth or dentures?	Functional limitation
Q 10	Have you had painful aching in your mouth?	Physical pain
Q 11	Have you had a sore jaw?	
Q 12	Have you had headaches because of problems with your jaws, teeth, mouth or dentures?	
Q 16	Have you found it uncomfortable to eat any foods because of problems with your jaws, teeth, mouth or dentures?	
Q 19	Have you been worried by jaw or dental problems?	Psychological discomfort
Q 20	Have you been self conscious because of your jaws, teeth, mouth or dentures?	
Q 21	Have jaw or dental problems made you miserable?	
Q 23	Have you felt tense because of problems with your jaws, teeth, mouth or dentures?	
Q 28	Have you had to avoid eating some foods because of problems with your jaws, teeth, mouth or dentures?	Physical disability
Q 32	Have you had to interrupt meals because of problems with your jaws, teeth, mouth or dentures?	
Q 33	Has your sleep been interrupted because of problems with your jaws, teeth, mouth or dentures?	Psychological disability
Q 34	Have you been upset because of problems with your jaws, teeth, mouth or dentures?	
Q 35	Have you found it difficult to relax because of problems with your jaws, teeth, mouth or dentures?	
Q 36	Have you felt depressed because of problems with your jaws, teeth, mouth or dentures?	
Q 37	Has your concentration been affected because of problems with your jaws, teeth, mouth or dentures?	
Q 40	Have you been less tolerant of your spouse or family because of problems with your jaws, teeth, mouth or dentures?	Social disability
Q 42	Have you been a bit irritable with other people because of problems with your jaws, teeth, mouth or dentures?	
Q 44	Have you felt that your general health has worsened because of problems with your jaws, teeth, mouth or dentures?	Handicap
Q 47	Have you felt that life in general was less satisfying because of problems with your jaws, teeth, mouth or dentures?	

6.5.4. Validity and reliability of OHIP and OHIP/TMD-I

This section reports findings about psychometric properties of OHIP and the proposed short form in an English speaking TMD population. It is essential that distilling the original OHIP does not compromise its psychometric properties.

6.5.4.1. Criterion validity

The correlation between full OHIP and OHIP/TMD-I was highly significant with a very strong correlation of 0.91 (Table 6.17).

6.5.4.2. Discriminant validity

A recent study that used OHIP in a TMD population found a mean difference of 0.53 per item between TMD sufferers and controls (John *et al.*, 2007). In the present study the mean difference per item between these two groups was 0.89 for the full OHIP and 1.53 for the OHIP/TMD-I. Clearly, the short specific version of OHIP reported here shows a much greater discriminant capability than that of the full OHIP.

Item selection criteria used in this study ensured that all identified items revealed a difference between patients and controls potentially sufficient to detect changes during treatment.

6.5.4.3. Convergent validity

Sixteen items of the MPI were included in this analysis, with their summary measures displayed in Table 6.15. On a scale of seven points (0 – 6) individual items' means ranged between 0.71 and 4.42 with an average of 2.28 per item (SE = 0.16). It should be noted, however, that higher scores do not necessarily imply worse QOL, as three items of MPI have their positive anchors on the right side, i.e. the higher the score, the better the QOL.

Pain scores, as measured by VAS1 (pain at the moment) and VAS2 (pain over the last month) were available from 108 patients. In both scales, values ranged between the two possible extremes, i.e. from 0 to 100. VAS 1's mean (SE.) were 36.37 (2.41), while VAS2's mean and SE were bigger: 51.30 (2.80) indicating that the pain on presentation was less than the pain experienced over the previous month.

Table 6.15. MPI summary measures for patients

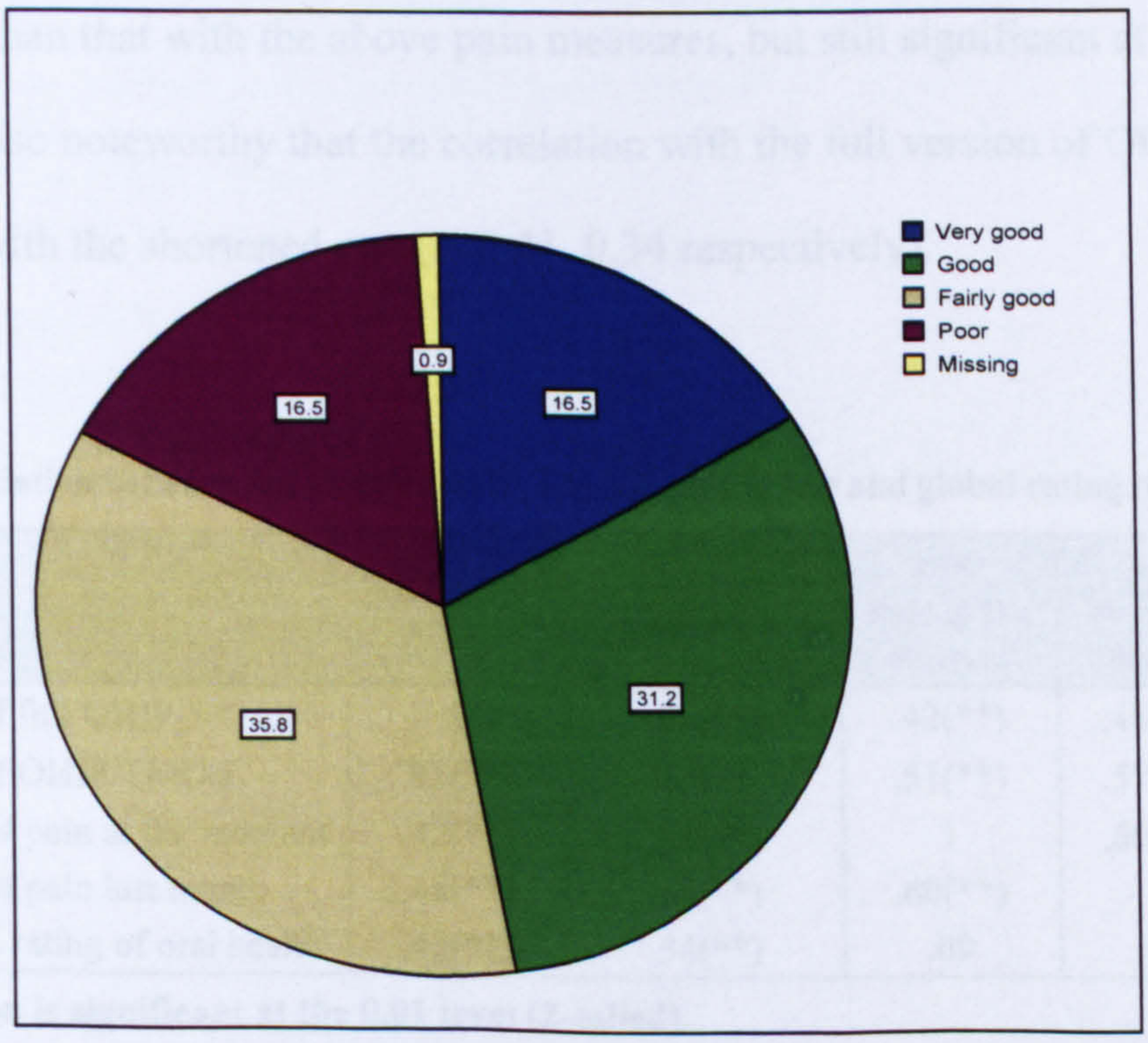
Item No.	MPI item*	Mean	Std. Error	Median
11	Control over life during the past month **	4.42	0.16	5
16	Ability to deal with problems during last month**	4.23	0.165	5
6	Overall mood during the past month**	3.36	0.138	3
7	Severity of pain during the last month	3.14	0.17	3
12	Suffering experienced because of pain	2.91	0.172	3
20	During the past month how tense or anxious have you been?	2.89	0.174	3
18	During the past month how irritable have you been?	2.72	0.176	3
1	Pain at the present moment	2.48	0.151	2
2	Pain problem interfere with day to day activities	1.97	0.163	2
4	Pain changed the amount of satisfaction	1.73	0.168	1
8	Pain changed ability to participate in recreational & social activities	1.43	0.165	1
3	Pain changed ability to work	1.34	0.16	0
9	Pain changed the amount of satisfaction from family-related activities	1.26	0.159	0
13	Pain changed marriage and family relationships	0.94	0.143	0
17	Pain changed ability to do household chores	0.94	0.152	0
19	Pain changed your friendships with people other than your family?	0.71	0.124	0

*** Items are sorted in a descending order according to their *Mean* score. Items 5, 10, 14, 15 are excluded.**

**** highlighted in orange are items with a positive scale, i.e. the higher the score, the better the QOL.**

The global rating of oral health (Figure 6.5) showed that, on a scale of five points (1- 5: *Excellent, Very good, Good, Fairly good, Poor*), patients rated their oral health as *fairly good* on average (median = 4). True values ranged between 2 and 5, and none reported their oral health as excellent.

Figure 6.5. Global rating of oral health



Correlation tests

A large correlation was found between sum of MPI and both the sum of full OHIP and sum of OHIP/TMD-I (r= 0.73) which was significant at the 0.01 level (Table 6.16).

Table 6.16. Correlation between MPI, full OHIP and OHIP/TMD-I

		Sum of MPI *	Sum of full OHIP	Sum of OHIP/TMD-I
Pearson Correlation	Sum of MPI	1	.73(**)	.73(**)
	Sum of full OHIP	.73(**)	1	.94(**)
	Sum of OHIP/TMD-I	.73(**)	.94(**)	1

*** Sum of MPI items after excluding missing data (items 5, 10,14,15 and case 31) and after reversing the positive scale in items 6,11,16.**
**** Correlation is significant at the 0.01 level (2-tailed).**

A medium correlation was found between sum of full OHIP and both VAS1 (pain at the moment) and VAS2 (pain last month) which was significant at the 0.01 level (Table 6.17). The correlation seemed to be slightly stronger with VAS2 (r =0.48) than with VAS1(r= 0.42). The correlation was improved for the OHIP/TMD-I score (0.57, 0.51 respectively). OHIP/TMD-I showed a slightly stronger correlation with VAS scales than did the full OHIP.

The correlation between global rating of oral health and both OHIP/TMD-I and full OHIP was weaker than that with the above pain measures, but still significant at the 0.01 level (Table 6.17). It is also noteworthy that the correlation with the full version of OHIP was stronger than that with the shortened one (r=0.41, 0.34 respectively).

Table 6.17. Correlation between full OHIP, OHIP/TMD-I, pain VASs and global rating of oral health

		Sum of full OHIP	Sum of OHIP/TMD-I	VAS1 Pain at the moment	VAS2 Pain last month	Global rating of oral health
Pearson Correlation	Sum of full OHIP	1	.91(**)	.42(**)	.48(**)	.41(**)
	Sum of OHIP/TMD-I	.91(**)	1	.51(**)	.57(**)	.34(**)
	VAS of pain at the moment	.42(**)	.51(**)	1	.60(**)	.09
	VAS of pain last month	.48(**)	.57(**)	.60(**)	1	.15
	Global rating of oral health	.41(**)	.34(**)	.09	.15	1

**** Correlation is significant at the 0.01 level (2-tailed).**

6.5.4.4. Stability (test-retest reliability)

The intra-class correlation coefficient ICC analysis showed that patients' scores on OHIP on both occasions were highly correlated. ICC measured 0.96 and 0.97 for full OHIP and OHIP/TMD-I respectively.

Wilcoxon test revealed a non-significant difference between the two measurements, both for the full OHIP and the short form OHIP/TMD-I (P = 0.1 and 0.07) respectively.

6.5.4.5. Internal consistency reliability

Cronbach's α coefficient was calculated to measure homogeneity within items of the full OHIP. This showed high internal consistency with a Cronbach' α of 0.95. Some items within each scale had, however, less correlation with other items in the scale. Four items in the full OHIP showed a Corrected Item-Total Correlation of less than 0.3 (Table 6.18). In other words, these items were measuring something different to the scale as a whole in this particular patient group, i.e. TMD. The total Cronbach's α coefficient of the full OHIP showed improvement if these four items were deleted (Table 6.18). Another two items, *Food catching in teeth* and *Sensitive teeth* showed improvement to the total Cronbach's α coefficient if deleted, but their Corrected Item-Total Correlation exceeded 0.325.

Cronbach's α measuring homogeneity within items of OHIP/TMD-I showed also high internal consistency (α = 0.94). All items within this short form had their Corrected Item-Total Correlation value greater than 0.44.

Table 6.18. Internal consistency for OHIP with low correlation

		Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Cronbach's α	Dentures have not been fitting properly	0.106	0.952
	Unable to eat with dentures	0.113	0.952
	Uncomfortable dentures	0.148	0.952
	Tooth doesn't look right	0.296	0.951

6.6. Discussion

The aim of this study was to develop a quality of life measure of TMD treatment outcomes. The Oral Health Impact Profile (OHIP) was used as a starting point and candidate items were identified as a basis for the intended TMD outcome measure.

6.6.1. Methodological issues

OHIP was chosen because it is a well-established oral health related quality of life (OHrQOL) measure, with a strong theoretical framework. It has been widely used in oral health research and proven to be effective as a discriminative tool (Slade, 1997b) and also as an outcome measure for different dental and oral interventions (Rogers *et al.*, 1998; Awad *et al.*, 2000; Allen and McMillan, 2003). However, it can be argued that OHIP's utility as an outcome measure for TMD patients might be questionable. This is because it was originally derived from interviews with a cohort of 64 patients aged 60 years old and over, hence the cohort might not be sufficiently representative for the general population, and specifically for the TMD population whose peak age group is mid 30s – mid 40s (LeResche, 1997b). The counter argument is that OHIP has been shown to be a useful OHrQOL in different age groups, for example in the U.K. National Dental Health surveys (Kelly, 2000). Furthermore, OHIP has been used successfully in TMD research (John *et al.*, 2002; Larsson *et al.*, 2004; Segù *et al.*, 2005; Wolfart *et al.*, 2005).

While this project aimed to identify OHIP-TMD relevant items from the existing index, ongoing work is taking place by another colleague to explore, through in-depth interviews with TMD patients, whether any other items may be needed to complement the short form OHIP. Given the ability of the index to measure OHrQOL in TMD patients and the structure of the index, many more items may not be required.

Although OHIP has been tested before on the TMD population in different studies, none of these studies was conducted using the English version of OHIP with one exception (Murray *et al.*, 1996). However, the authors used only part of OHIP and excluded 21 items based seemingly on their subjective judgement. The present study was the first to systematically distil OHIP and eliminate redundant items to improve its performance as a potential TMD treatment outcome.

Shortening OHIP to a more condition (TMD)-specific subset of items would be very helpful, firstly to improve its precision and consequently its psychometric properties (Allen and Locker, 2002; Allen, 2003). Secondly this would reduce the respondent burden and make the questionnaire easier to complete, which would consequently improve the response rate (Locker and Allen, 2002). Thirdly, the short form would be easier to use in clinical settings and less costly in large national surveys.

The first step in this research involved adapting OHIP to be more suitable for TMD conditions, and consequently to improve the psychometric properties of the intended short version. This included re-wording OHIP items so it included a reference to the *Jaw*. The need for adding this word was previously recognised by the authors of the Swedish version of OHIP (Larsson *et al.*, 2004).

The reference period *(recall period)* in this study was changed to one month since the one-year period used in the original OHIP was not appropriate for TMD, where the condition is often unstable and a shorter reference period is needed. The choice of one month was based on the results of the first two studies, and has also been supported by other authors (John *et al.*, 2002).

A case series design was used in this study. The sample size was initially determined at 99 which would allow a precision of percentage estimate for each item within 10 percentage points, and would also be more than sufficient to show the expected score change with

treatment. This sample size exceeded the size needed for a study power of 90% to just compare between patients and controls, but was necessary for future analyses on items' sensitivity to changes with treatment. This is because a large sample is needed to detect small differences in the condition and where drop out from the study is anticipated and needs to be anticipated. This sample size was later extended to 110 patients in order to examine the stability (test-retest reliability) of OHIP.

The patient sample covered a large range of age groups and consequently was similar to epidemiologic research on TMD (Dworkin *et al.*, 1990; LeResche, 1997b). However, subjects below the age of 18 years were not recruited to comply with the legal-ethical considerations related to consent. The wide sample made it possible to include a wide spectrum of TMD diagnoses, which ensured representation of different conditions with the potential for various complaints and impacts on quality of life. Patients participating in this study underwent a variety of treatments prescribed on the basis of what was considered most appropriate by the responsible consultant.

A control group was necessary to eliminate items that appear to have high impact on patients' QOL but are in fact representing "background noise", in other words, items that score frequently irrespective of TMD status. Ideally, the control group should have been the normal population. Unfortunately, however, no normative values for an English speaking population have been established. The best available work was the National Dental Health Survey in the United Kingdom 1998 (Kelly, 2000), which used OHIP-14. A survey that used the full OHIP in a random sample of community dwelling Canadians is also available (Locker and Slade, 1993). However, participants were all 50 years old and older who cannot be age matched to the TMD population whose peak age group is mid 30s – mid 40s (LeResche, 1997b). For this reason, OHIP was administered to an age and sex matched TMD-free control group.

6.6.2. Differences between TMD patient and control groups

In order to examine the difference between the two case groups, data were first explored for normality of distribution. Individual item scores for both patients and controls were all skewed. This finding is in accordance with Slade (2002) who noted many respondents reporting no impact at the defined threshold, resulting in summary scores being non normally distributed and restricting the use of parametric statistics (Slade, 2002). For this reason and because the data were at the ordinal level (Bowling, 2002), non parametric tests were used to compare between the two groups.

As shown in section 6.5.2.1, overall, OHIP was clearly able to discriminate between the two groups and to reflect the impact of TMD on patients' quality of life. Some items showed no statistical difference between the two groups. These are not anticipated to capture the expected smaller difference between patients before and after treatment. These, however, were only eight items, which were considered redundant and were later eliminated. The relatively large number of items (41) showing statistically significant difference between patients and controls reflects that TMD conditions result in an impact on aspects of life other than one might anticipate. This issue has been investigated further (Steele *et al.*, 2006) but it is beyond the scope of this thesis.

As the statistically significant difference was not enough to identify the items most relevant to TMD, I measured the difference in each item between patients and controls using four analyses: *Mean*, *Median*, *Prevalence* (%FOVO) and *Item-Impact*. These four analyses allowed the data to be viewed from different aspects.

Using *Means* is quite a useful approach to analyse the data despite the statistical shortcoming that the data were not normally distributed. In addition they were also at the ordinal level which cannot be averaged or mathematically manipulated (Bowling, 2002). In other words, a mean cannot be calculated for "occasionally" and "fairly often" and so obtain a

new rate. However, the values attributed to these levels of frequency, i.e. their orders in the scale can be averaged (Slade, 2002). The resulting value does not have a clear narrative meaning, but can be used to compare groups (Slade, 2002). For example, if "occasionally" with "fairly often" are given the values 2 and 3 respectively, we cannot interpret the meaning of a value of 2.7, but we can say that a group whose score is 2.7 is suffering more frequently than another with score 2.5.

The *Item-Impact* method used in this project was the modified version of *Item-Impact* that has been suggested by Allen and Locker (2002). However, there were three points of dissimilarity. First, unlike Allen and Locker (2002), I did not rely merely on patient's item-impact values, but on the differences between these values and those of the controls. Nonetheless, I found only trivial difference between order of items by "patients' preference", and that according to difference between the two groups. Allen and Locker (2002) also used the proportion of patients who reported any degree of impact from hardly ever to very often. Instead, I opted to report the proportion of patients whose lives are compromised on a constant or frequent basis by using a more stringent cut-off point for the responses (fairly or very often – FOVO). Finally, I did not consider items' weights in calculating item impacts. This is because OHIP item weights did not have any effect on the final results (Allen and Locker, 1997). In a different study, John *et al.* (2002) found a strong correlation ($r_s = 0.99$) between the sum of weighted and unweighted data and preferred the use of simpler unweighted score. I followed this line in this study.

Differences in summary scores between patients and controls

All individual item summary scores varied between the two case groups. For example, a difference of 0.89 unit was noted in the mean item score. Also, for all but two items in the control group, median scores were 0 reflecting no impact on quality of life. By contrast, median scores for 26 items in the patient group were equal or greater than one unit.

Furthermore, the percentage of items reported as occurring fairly or very often (%FOVO) was ten times higher in the patient group than the control group (20% : 2%). Finally, on average the item-impact score for patients was 25 times that of the controls (35 : 1.5). These differences clearly show that the English version of OHI P was able to discriminate between patients and controls.

What was also clear from the results is that items varied extensively in their summary scores within the same group. In particular, some items were highly prevalent amongst sufferers, while others less frequently affected quality of life.

These findings were similar to those reported by John *et al.* (2002) and Larsson *et al.* (2004) who used the full OHIP in TMD patient samples in Germany and Sweden respectively. For example, mean overall OHIP score in the patient sample was 60.5 (SD=31.6). The German values were 45 (30.9) and the Swedish values were 61 (39).

It was difficult to make detailed comparisons between the present findings and previous studies that have used OHIP in TMD due to differences in the items and analyses used (Murray *et al.*, 1996; Segù *et al.*, 2005). Furthermore, Murray *et al.* omitted in the OHIP questions references to problems with mouth, teeth or dentures and limited the cause of the impact on quality of life to pain. However, within these limitations I could conduct some comparisons which are reported in their relevant sections.

Variation of TMD impact on OHIP domains

The present data suggest that the psychosocial impact of TMD is considerable. For example, the *psychological discomfort* and *psychological disability* were the domains with largest differences between patients and controls in mean scores (*Mean per item*). The *physical pain* domain came third in order. This might be due to the fact that this domain comprises many less relevant items to TMD such as gum pain and toothache, which might have compromised

the domain score. On the other side of the scale, I found that *social disability* and *handicap* domains showed the lowest scores in the patient group. This might be interpreted by the fact that advanced impact on quality of life is uncommon and there is a floor effect in these two domains. This also explains why *social disability* and *handicap* domains were not represented in the first round of identifying TMD relevant items.

The percentage differences in reporting items at the FOVO threshold between patients and controls were striking. TMD sufferers reported frequent or constant impact on different domains of quality of life that was 4-20 times higher than those of the controls (Table 6.8).

These findings are in close agreement with Larsson *et al.* (2004), Murray *et al.* (1996) and Segù *et al.* (2005). It was not possible, however, to compare the present findings about the *pain* and *function* domains against the latter two studies who initially excluded the entire *pain* domain and many items from the *function* domain from their version of OHIP. Also, the percentage of patients reporting a frequent or constant impact in Murray *et al.*'s study (1996) was higher than all other studies including the present findings. Although the domain order was almost identical, some domain scores in Murray *et al.*'s were more than twice as high as our data. For example, the psychological discomfort impact was reported by 33% of the present patient sample compared with 667.8% in Murray *et al.*'s study. This unusual high level of suffering was explained by the authors as a result of recruiting patients suffering neurological pain disorders who scored higher on psychological related domains than did TMD patients, hence increasing the total domain score. Furthermore, these authors recognised the fact that their patient sample was drawn from a tertiary care clinic, whose patients presented with the more severe conditions.

6.6.3. Candidate items (OHIP/TMD-I)

As the statistical difference between patients and controls was not sufficient on its own to exclude all irrelevant items, another method was needed to distil the original OHIP. Different methods to shorten long instruments have been used in the past such as regression factor and item-impact analyses. Recommended uses of these methods were discussed in the literature review section in greater detail (section 2.4.3.5). Based on that review I opted to select items based on their item-impact score. While this is the method of choice to shorten long evaluative questionnaires (Guyatt *et al.*, 1986; Juniper *et al.*, 1996; Locker and Allen, 2002), the method ranks items, but does not yield cut-off points between relevant and less relevant items, nor does it produce meaningful units that can be used to draw such points. Therefore, other criteria were needed to "filter" OHIP items and identify TMD relevant items in a systematic way.

The criteria used dictated that items with the largest difference between patients and controls in *Item-impact* will be included if they showed statistically significant differences and demonstrated clinically meaningful differences in the *Mean* and *Median* scores between TMD sufferers and controls. Moreover, in order to maintain the structure of the index, additional items were chosen to represent all OHIP domains. Ideally, the next step would be checking that all items are sensitive to change with treatment. Follow-up data were collected to carry out this analysis, but this constitutes further work and is not reported in this thesis.

As expected, the *Item-Impact* analysis produced a continuum of values and did not show any cut-off points that could be used to discriminate between high- and low-impact items. It was found that all four summary scores (*Mean*, *Mean*, *Prevalence*, and *Item-impact*) consistently identified the same top items.

The concept of a clinically meaningful difference (patient/control) was adopted. Items that reached the determined threshold were chosen initially as candidate items. The minimum

difference of OHIP score between the groups was taken as one unit in the mean item score. Although this was arbitrary it allows an overall narrative difference that is meaningful in clinical research. If OHIP fails to demonstrate a minimum difference of one unit per item, it is unlikely that it will be able to capture the smaller difference between patients' scores before and after treatment. In other words, it will not be able to demonstrate improvement with treatment in a clear clinical term – the very reason for which it was shortened.

At this stage, 16 items were identified in the *Mean* analysis, of which 14 items were top items in the *Median* analysis. Also, amongst these 16 items were the 15 most prevalent items (%FOVO), and the top 15 in the *Item-Impact* analysis.

The 16th item identified by the *Mean* test was *"question 36: felt depressed because of problems with jaws etc..."* which was not amongst top items in the *Median* analysis nor did it exceed the arbitrary threshold of 20% difference in FOVO. The decision to include or exclude this item was based on the argument that shortening long questionnaires always involves a combination of statistical considerations and subjective judgement (Coste *et al.*, 1997; Juniper *et al.*, 1997). Pure statistical methods may, for instance, exclude items with low prevalence, although these may be rated high by a minority of patients. Therefore, a clinical decision is needed to ensure that this minority is represented. This was the case for question 36. It was felt that excluding this item from the intended questionnaire would omit an improvement outcome for 16% of patients who reported being depressed often or very often (versus 1% only in the control group). For that reason, item 36 was included for the time-being. This identified 16 items in the intended instrument.

Maintaining OHIP structure

Having established OHIP's main TMD related items, their distribution was checked within the OHIP domain framework. It was found that the domains reflecting the most severe impact

on quality of life, i.e. *Social disability* and *Handicap* were not represented. This is explained by the fact that only a small proportion of the TMD population reported impacts on their lives at this level.

In order to maintain OHIP's theoretical hierarchal structure and still include items that are most important to patients, it was necessary to initially include at least the top two items in the *Item-Impact* analysis from each domain. This was important for two reasons. Firstly, to include items affecting a minority of TMD patients who suffer the more extreme impact on quality of life, e.g. related *Social disability* and *Handicap*. This argument is in line with findings by Segù *et al.* (2005) who reported a significant difference in the above domains between patients and controls in a subgroup of the patient group diagnosed as having an advanced condition. Similarly, Murray *et al.* (1996) found that around 40% of the orofacial patient population (including patients with TMD) reported the above domains at the fairly or very often (FOVO) level. Secondly, Moran *et al.* (2001) argued that when reducing an instrument at least two items per domain should be retained in order to maintain responsiveness and reliability at an acceptable level. This argument has been supported by previous work to shorten OHIP (Slade, 1997b) where the author maintained two items per domain to produce OHIP-14.

This analysis identified 20 items for OHIP/TMD-I with each domain being represented by two or more items, except *Functional limitation* with one item only. Further work aiming to investigate the need to add items to OHIP to improve its content validity may be able to cover the shortage of this domain. Also, further work is needed to exclude items with low responsiveness to changes with treatment, which would compromise the performance of the proposed TMD quality of life measure.

Table 6.19 compares candidate items identified by this study as a TMD-specific short form of OHIP and those chosen by Murray *et al.* (1996). Of the 30 items chosen by the latter

authors, 16 were identified in this study. The remaining 14 items were either not included in the original OHIP (2 items), related to speech (3), to appearance (1), to tooth brushing (1), to social impact (4) or to general disability (3). While the last two categories were represented in the present study, other items were clearly not related to TMD. On the other hand, four items that exist in the present study were excluded by Murray *et al.* (1996). These were related to pain or to general health.

Table 6.19. Comparison between candidate items and items chosen by Murray *et al.* (1996)

OHIP item	Murray *et al.*	Moufti *et al.*
01. Difficulty chewing any foods	1	1
16. Uncomfortable to eat any foods	1	1
19. Worried by jaw or dental problems	1	1
20. Being self conscious	1	1
21. Being miserable	1	1
23. Feeling tense	1	1
28. Avoid eating some foods	1	1
32. Having to interrupt meals	1	1
33. Sleep being interrupted	1	1
34. Being upset	1	1
35. Difficult to relax	1	1
36. Feeling depressed	1	1
37. Concentration being affected	1	1
40. Less tolerance of spouse or family	1	1
42. Bit irritable with other people	1	1
47. Life in general less satisfying	1	1
02. Trouble pronouncing any words	1	0
24. Speech being unclear	1	0
25. People misunderstood some words	1	0
27. Being unable to brush teeth properly	1	0
31. Avoid smiling	1	0
38. Being a bit embarrassed	1	0
39. Avoid going out	1	0
41. Trouble getting on with other people	1	0
43. Difficulty doing your usual jobs	1	0
46. Unable to enjoy other people's company as much	1	0
48. Totally unable to function	1	0
49. Unable to work to full capacity	1	0
10. Painful aching in mouth	0	1
11. Having a sore jaw	0	1
12. Having headaches	0	1
44. General health has worsened	0	1
* *Avoid eating with other people*	1	N/A
* *Longer to complete a meal*	1	N/A

*** Items added to original OHIP by Murray *et al.* (1996)**
0 = item is not in this version, 1 = item is in this version

6.6.4. Validity and reliability of OHIP and OHIP/TMD-I

It has been argued that reducing the number of items may adversely affect measurement. This is especially the case with the content validity and reliability (Nunnally and Berbstein, 1994; Locker and Allen, 2002). It is therefore essential to test the psychometric properties of a newly derived short form even if the mother instrument has been validated. It is highly recommended that validation of the new short version should be conducted in a sample other than the one it was derived from (Coste *et al.*, 1997; Locker and Allen, 2002).

What is reported in this study is not the validation process of a short form of OHIP, e.g. OHIP/TMD-I, but an examination of OHIP-49's psychometric properties in an English-speaking TMD population and an exploration of how these properties may change after excluding redundant items.

Content and criterion validity

There has been debate in the literature about the content validity of the full OHIP. Although the full version was proven to be a valid instrument for cross-sectional and longitudinal studies on TMD (John *et al.*, 2002), other studies opted to administer only part of OHIP, using items that were believed to be more TMD relevant (Murray *et al.*, 1996; Segù *et al.*, 2005). These authors recognised the burden that non TMD related items might have with consequences on the instrument's performance.

In this study, it was found that the full OHIP was able to capture the impact of TMD on quality of life. However, by examining the performance of each item of OHIP, items with weak performance were identified and excluded, thus enhancing the content validity of OHIP. This was further complemented by amending the reference period to one month, and adding the word "jaw" to the source of patients suffering in the body of OHIP items. In addition, the short version contained items from all OHIP domains in order to maintain the hierarchal structure of OHIP (i.e. Locker's model), thus agreeing with Slade (1997b) and with Allen and Locker (2002).

Further development of the TMD outcome measure will involve examining whether other items need to be added to OHIP. Murray *et al.* (1996) and John *et al.* (2002) suggested the addition of two items: *"Taking longer to complete a meal"* and *"Avoiding eating with others"*. A parallel research involving in-depth interviews with TMD patients is being undertaken by the same research team to address this issue. But, it is beyond the scope of this thesis to report the findings of that research.

Criterion validity for OHIP in the TMD population has been established (John *et al.*, 2002) (Murray *et al.*, 1996; Segù *et al.*, 2005). The full version works as the gold standard against which the validity of the short versions can be examined (Slade, 2002). In this study, a very strong correlation of 0.91 was found between the full OHIP and OHIP/TMD-I reflecting the validity of the short form.

Construct validity

This aspect of validity has two components: the *Discriminant validity* (also known as *Known-group validity* or the *Group validity*) and the *Convergent validity*.

The discriminant validity was explored by comparing OHIP scores in the TMD patient and control groups. This demonstrated the ability of full OHIP to discriminate between patients and controls; not only 41 of its 49 items showed statistically significant difference between the two groups, but also a difference of 0.89 unit per item was recorded on average. The proposed subset of items, OHIP/TMD-I, poses a stronger ability to discriminate between the two groups. This was demonstrated by the fact that all its items have statistically significant differences between the two groups, and that a difference of 1.53 unit per item was recorded on average.

To test the convergent validity of OHIP, different parallel patient-derived measures were used, including perceived pain severity and its impact on quality of life and global rating of mouth health. Use of these parallel measures was necessary not only to explore the convergent

validity of OHIP in the TMD population, but also to examine OHIP's sensitivity to changes with treatment in a future work.

The need for different parallel instruments to check OHIP against came from the fact that OHIP is a multidimensional instrument. Therefore different measures were needed to explore the convergent validity of the different OHIP domains. This same argument was used before to validate the Swedish version of OHIP (Larsson *et al.*, 2004). Moreover, TMD itself is a multidimensional condition. Previous studies (Segù *et al.*, 2005) validated OHIP against measures of pain only arguing that the greater the pain, the greater is the functional limitation and psychological discomfort. While pain is a major source of suffering, it cannot be considered the only source. For example, in the second part of this project, it was reported, that the worries from the condition was sometimes the reason for seeking the medical attention, and many patients reported improvement in their conditions after a consultation session.

The first of these self-report measures was the Multidimensional Pain Inventory MPI (Kerns *et al.*, 1985). MPI is considered "the gold standard multi-axial pain self-report instrument" (Newton-John, 2002). Furthermore, it is recommended as one of the best two indices in pain trials (Turk *et al.*, 2006). In this study only part one of the MPI was used. This is because part two of (about relatives' support) and part three (about the impact on daily activities) are more related to severe or generalized pain conditions where others' support and performing daily activities are key issues for the patients. Furthermore, part one has items that cover these two parts.

The second self-report measures were two visual analogues scales of perceived pain severity. Perceived pain scales have been extensively used to validate long self-report measures (Locker and Miller, 1994; Locker, 1998), and different types of these scales have been used. Numerical Rating Scales (NRS), for example, have been used to explore the validity of OHIP (Murray *et al.*, 1996; Segù *et al.*, 2005). Although this is technically less

demanding than visual analogue scales consisting of continuous lines on which patients mark the extent of their pain, the VAS is preferred as it avoids leading respondents to clustered answers (see section 2.4.4.4).

The third patient-derived measure used to check OHIP's convergent validity was a global rating of oral health. This has been used to validate both the original and the German version of OHIP (Locker and Slade, 1993; John *et al.*, 2002). Other workers have used different scales of perceived health, For example, to validate the Swedish version of OHIP the authors asked patients to rate their general health but not the health of their mouth (Larsson *et al.*, 2004).

OHIP showed good convergent construct validity in the TMD population. All tests of correlation between either the full OHIP or the subset of items, i.e. OHIP/TMD-I and other pain self-report measures showed large correlation which was significant at the 0.01 level. The correlation with perceived oral health was weaker than that with pain measures and weaker in OHIP/TMD-I but still significant at the 0.01 level. These findings are comparable to findings in the Swedish OHIP (medium correlation with self-report pain measures) (Segù *et al.*, 2005), and the German OHIP (medium correlation with global rating of oral health) (John *et al.*, 2002).

Reliability

The test-retest reliability for established instruments must be re-examined when they are used in a different population than that on which they were originally tested (Litwin, 1995). But testing stability is challenging in clinical settings because the condition itself is often unsettled. The interval between the two occasions has to be long enough to avoid memory effects, i.e. patients recalling their initial answers, and not too long so that the condition changes.

In this study, both the full OHIP and the short form OHIP/TMD-I showed high stability, ICC measured 0.96 and 0.97 respectively. These figures are comparable with those for the

Swedish OHIP where stability was measured over one to two weeks in a sample of TMD patients (ICC for OHIP's different domains measured between 0.87 and 0.98).

The internal consistency reliability was very high for both the full OHIP (0.95) and the proposed short form (0.94). Cronbach's α coefficient showed that the exclusion of four items from OHIP would improve the internal reliability. These four items (related to denture and food catching) had been previously shown not to have any statistical significance differences between patients and controls, and none of these was in the OHIP/TMD-I short form.

These findings are consistent with previous studies, e.g. Larsson *et al.* (2004) who found am internal reliability of 0.91.

6.6.5. Strengths and limitations

This study was the first to systematically distil OHIP to produce a TMD specific version that might be suitable, once fully tested, as a TMD treatment outcome. Also, although the English version of OHIP has been previously administered to TMD patients (Murray *et al.*, 1996), the authors used only part of the original instrument. The present study was the first to use the full English version in a TMD population and to explore in detail the impact of TMD on patients' lives. The use of the full version showed that items not expected to show higher impact than that seen in TMD free controls have indeed scored higher. These findings will undergo a future exploration.

Two aspects of the protocol were adopted to give more strength to this study: First, the sample size allowed for a study power of greater than 90% and for a power of 80% for the future sensitivity anlysis. A wide range of TMD diagnoses were represented. All patients completed baseline OHIP questionnaires before any consultation from the clinical staff in order to avoid bias.

If the aim of this project was to validate OHIP or its newly derived short form I should have used additional parallel measures to ensure the convergent validity of all OHIP's subscales. This includes, in addition to the global rating of health and measures of pain, measures of the jaw/mouth function such as the Mandibular Function Impairment Questionnaire (MFIQ) (Stegenga *et al.*, 1993), and a measure of the psychological health such as RDC / Axis-II (Dworkin and LeResche, 1992). Similarly, the small patient sample used for the test-retest analysis does not permit for generalising conclusions about OHIP's stability other than for the purpose of initial exploration of psychometric properties.

The type and length of treatments provided in this study were not standardised, but rather based on what was considered most appropriate by the responsible consultant. This was deemed necessary to represent a wide spectrum of treatment modalities, and to avoid influencing the consultant's opinion.

The candidate items were derived using the item-impact method and the ranking of mean and median values. Other methods to shorten long instruments are available such as the *Factor analysis*, which may be able to show how items cluster. This is turn may inform the theoretical structure of OHIP. In this respect, John *et al* (2004) have postulated that OHIP should be reconfigured into four rather than seven domains. These were: psychosocial impact, orofacial pain, oral function and appearance. The heterogeneous nature of TMD dictates a large number of subjects for this type of analysis. Hence, it would be premature to apply it to the current data set.

6.7. Conclusions

1. *(Objective1)* – OHIP-49 was adapted to improve its utility as a measure of TMD treatment outcome. The word *"jaw"* was added to the question wording as a source of suffering, and the reference period was altered to one month.
2. *(Objective 2)* – The adapted version of OHIP was administered on 110 TMD patients and matching TMD free controls.
3. *(Objective 3)* – There are profound differences between TMD patients and healthy controls in OHIP scores. OHIP-49 is capable of discriminating between TMD patients and TMD free subjects. This is a prerequisite for the instrument to be able to demonstrate difference between pre- and post-treatment. The psychosocial impact of TMD on patients is substantial. The results of this work will help clinicians dealing with TMD to address patients' suffering in their treatment plan.
4. *(Objective 3)* – All item scores were higher for patients than controls, but the difference was not always statistically or clinically significant. Redundant items may reduce the performance of OHIP, and need to be eliminated to make a better instrument.
5. *(Objective 4)* – Twenty candidate items which may form the basis of a clinical outcome measure were identified and called "OHIP/TMD-I". First, items showing no statistical significance difference between TMD patients and controls were eliminated. Second, items showing largest differences between TMD patients and controls were identified using summary measures including Mean, Median, Prevalence and Item-Impact. Third, items were selected taking into consideration the theoretical framework of OHIP domains.
6. *(Objective 4)* – The domain functional disability was only represented by a single item and may require further items outwith OHIP
7. *(Objective 5)* – The psychometric properties of the full OHIP and of the short form "OHIP/TMD-I" were initially explored including content, discriminant, convergent and criterion validity; test-retest and internal consistency reliability. Both versions showed satisfactory properties, with the short form performing better in some aspects. This cannot be generalised before testing the short form in a new sample.

6.8. Summary

The aim of this study was to start the development of a quality of life outcome measure for TMD. The intention was to obtain a TMD specific short form of the Oral Health Impact Profile (OHIP) and to explore aspects of psychometric properties for the derived instrument, the so-called "OHIP/TMD-I".

An adapted version of OHIP was used. Adaptation included adding the word *"jaw"* to the question wording and changing the reference period in the original OHIP to one month. A prerequisite for this version to be a useful outcome instrument is to demonstrate ability to discriminate between TMD patients and TMD free controls.

TMD patients completed the adapted version with the aim of identifying OHIP items with the largest impact on patients' quality of life in comparison to matching controls. Criteria were set to select candidate items for the short form, and 20 items were identified. The validity and reliability of the full OHIP and its short form were examined by a series of tests.

The adapted English-language version of OHIP was able to discriminate between patients and controls. The identified 20 items had comparable or better psychometric properties than those of the original OHIP. Further work is needed to examine the sensitivity to change of these items, an essential characteristic in outcome measures, and to investigate whether items are missing from OHIP and needed to be added. Once fully completed, the psychometric properties of the short form need to be examined in a new independent patient sample.

Chapter 7.
Conclusion of project and further work

7.1. Conclusions

Study One established optimal review and reference periods for TMD outcome assessment by examining pain recovery progress of patients undergoing conservative treatment for TMD.

Part A of this study identified patients showing definite improvement to treatment using two methods. First, by visually analysing pain / time plots of daily pain intensity, and second by examining these pain scores numerically. The visual examination revealed patients had three patterns of pain / time response characteristics: Improvers, non-improvers and borderline improvers. The numerical analyses identified almost the same patients identified by the visual assessment as improvers. The best agreement occurred with a combined 50% reduction in mean and maximum monthly pain scores (aggregated sensitivity and specificity = 1.74). This was defined as a suitable cut-off point to discriminate numerically between improvers and non-improvers.

Optimal review period and reference periods were established in part B of this study. These were 15 weeks and 4 weeks respectively.

Study Two was conducted to triangulate with *Study One*. It used qualitative data from ten in-depth interviews with patients undergoing conservative treatment for TMD and showing improvement. The study found that perception of suffering from TMD and its recovery were mainly linked to the experience of pain. This confirmed that the approach adopted in *Study One*, i.e. establishing the review and reference periods based on pain recovery was a reasonable approach.

The interviewees reported temporal characteristics of recovery such as gradual and fluctuating improvement similar to what was found in *Study One*. However, the limitations of qualitative approach did not allow for an accurate estimation of the review and reference periods. Nevertheless, within these limitations time frames for these periods suggested in both studies were consistent.

In *Study Three* a TMD-specific short form of OHIP was derived and aspects of its psychometric properties were explored. This subset of OHIP items comprised the core of a TMD quality of life outcome measure under development.

The Oral Health Impact Profile (OHIP) was adapted to improve its utility and content validity for TMD. The adapted version included the word *"jaw"* as a source of the impact on quality of life and had a one-month reference period.

The adapted version demonstrated remarkable ability to discriminate between TMD patients and TMD-free controls, a prerequisite to be a useful as an outcome instrument. The discriminative validity was further improved by reducing OHIP to a 20-item short version, OHIP/TMD-I, maintaining items with significant statistical and clinical differences between patients and controls. OHIP/TMD-I performed comparably or better than the full OHIP in other psychometric tests. Further work needed to continue the development of the proposed TMD outcome instrument is outlined below.

7.2. Recommendations for further investigation

7.2.1. Study One

1. The pain diaries collected in this study offered valuable information that could be used for numerous future analyses, including:

   a. Investigating the role of daily psychological stress on the illness status. This could be explored by analyzing the differences in jaw pain level, headache hours and consumed number of painkillers with relation to weekdays versus weekends.

   b. Investigating the relationship between jaw pain and headache, and exploring the initial reason for consuming painkillers in TMD: jaw pain versus headache.

   c. Testing the validity of summary measures of pain intensity taken at follow-up appointments against daily data, and trying to explore what patients refer to by summary measure: average, maximum, or recent experience.

2. Standardised visual assessment of plotted pain changes with treatment is a promising method of outcome assessment in clinical trials. While characteristics of pain patterns may differ across chronic conditions, the same principle is still applicable.

3. Whilst systematic, and preferable to the use of an unsubstantiated statistical approach alone, the visual assessment is still clinician-led. Further work may be required to establish the nature of a clinically meaningful change, from the patients' perspective as well as the clinicians (e.g. rescue medicine).

4. Analysing rescue medicine data has potentially an important value in assessing TMD treatment outcome, especially if analgesic type and use is standardised.

5. Area under the curve is a complicated test, which is not available in all statistics packages. Assuming only few data are missing, it is not favoured over the simple Mean test.

7.2.2. Study Two

1. This study revealed interesting information about the suffering of TMD and patients' journey through illness and treatment. In-depth qualitative research is needed to give further insight on the lived experience of TMD and orofacial pain in general. This seems necessary to ensure that management and outcome assessment of these conditions map appropriately to patients' own perceptions and expectations. In this context, this further insight can also help to add any themes to OHIP that may be deemed necessary. Another research project is currently undertaken by the same research team to address this point and also to develop a theoretical model for TMD.
2. Different characteristics of TMD pain and symptoms were expressed by the participants in this study. These are worthy of further exploration, which could potentially improve our understanding and differential diagnosis of TMD.

7.2.3. Study Three

Further work is needed to:

1. Explore if extra items needed for OHIP/TMD-I using qualitative interviews with patients. This is ongoing work by the research team.
2. Analyse the follow-up data for the TMD patient group to determine sensitivity of the OHIP/TMD-I items to change with treatment
3. Examine the psychometric properties of the final set of items using an independent sample of TMD patients
4. Explore the difference in TMD impact on patients with different subgroups, including age, gender and RDC/TMD diagnosis sub-groups
5. Check the characteristics of the control groups against the British National Survey (Kelly, 2000)
6. Perform a factor analysis to verify the appropriateness of the domain structure as it relates to TMD

Bibliography

Aaron LA, Burke MM, Buchwald D (2000). Overlapping conditions among patients with chronic fatigue syndrome, fibromyalgia, and temporomandibular disorder. *Archives of Internal Medicine* 160:221-227.

Aaron LA, Mancl L, Turner JA, Sawchuk CN, Klein KM (2004). Reasons for missing interviews in the daily electronic assessment of pain, mood, and stress. *Pain* 109:389-398.

Aaron LA, Turner JA, Mancl L, Brister H, Sawchuk CN (2005). Electronic diary assessment of pain-related variables: Is reactivity a problem? *Journal of Pain* 6:107-115.

Adulyanon S, Vourapukjaru J, Sheiham A (1996). Oral impacts affecting daily performance in a low dental disease Thai population. *Community Dent Oral Epidemiol* 24:385-389.

Aggarwal VR, Lunt M, Zakrzewska JM, Macfarlane GJ, Macfarlane TV (2005). Development and validation of the Manchester orofacial pain disability scale. *Community Dentistry and Oral Epidemiology* 33:141-149.

Al-Ani MZ, Davies SJ, Gray RJ, Sloan P, Glenny AM (2004). Stabilisation splint therapy for temporomandibular pain dysfunction syndrome. *Cochrane Database Syst Rev*:CD002778.

Allen PF, Locker D (1997). Do item weights matter? An assessment using the oral health impact profile. *Community Dental Health* 14:133-138.

Allen PF, McMillan AS, Walshaw D, Locker D (1999). A comparison of the validity of generic- and disease-specific measures in the assessment of oral health-related quality of life. *Community Dent Oral Epidemiol* 27:344-352.

Allen PF, McMillan AS, Locker D (2001). An assessment of sensitivity to change of the Oral Health Impact Profile in a clinical trial. *Community Dent Oral Epidemiol* 29:175-182.

Allen PF, Locker D (2002). A Modified Short Version of the Oral Health Impact Profile for Assessing Health-Related Quality of Life in Edentulous Adults. *International Journal of Prosthodontics* 15:446-450.

Allen PF (2003). Assessment of oral health related quality of life. *Health Qual Life Outcomes* 1:40.

Allen PF, McMillan AS (2003). A longitudinal study of quality of life outcomes in older adults requesting implant prostheses and complete removable dentures. *Clinical Oral Implants Research* 14:173-179.

Allen PF, Thomason JM, Jepson NJ, Nohl F, Smith DG, Ellis J (2006). A randomized controlled trial of implant-retained mandibular overdentures. *J Dent Res* 85:547-551.

Allison P, Locker D, Jokovic A, Slade G (1999). A Cross-cultural Study of Oral Health Values. *Journal of Dental Research* 78:643-649.

Altman DG (1991). Practical statistics for medical research. London: Chapman and Hall.

Anderson R, Thomas DW (2003). 'Toothache stories': A qualitative investigation of why and how people seek emergency dental care. *Community Dental Health* 20:106-111.

Anderson R (2004). Patient expectations of emergency dental services: A qualitative interview study. *British Dental Journal* 197:331-334.

Atchison KA, Dolan TA (1990). Development of the Geriatric Oral Health Assessment Index. *Journal of dental education* 54:680-687.

Auerbach SM, Laskin DM, Frantsve LME, Orr T (2001). Depression, pain, exposure to stressful life events, and long-term outcomes in temporomandibular disorder patients. *Journal of Oral and Maxillofacial Surgery* 59:628-633.

Awad MA, Locker D, Korner-Bitensky N, Feine JS (2000). Measuring the effect of intra-oral implant rehabilitation on health-related quality of life in a randomized controlled clinical trial. *J Dent Res* 79:1659-1663.

Bagewitz IC, Söderfeldt B, Nilner K, Palmqvist S (2005). Dimensions of oral health-related quality of life in an adult Swedish population. *Acta Odontologica Scandinavica* 63:353-360.

Baker SR, Pankhurst CL, Robinson PG (2006). Utility of two oral health-related quality-of-life measures in patients with xerostomia. *Community Dent Oral Epidemiol* 34:351-362.

Barbour RS (1999). The case for combining qualitative and quantitative approaches in health services research. *Journal of Health Services Research and Policy* 4:39-43.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961). An inventory for measuring depression. *Arch Gen Psychiatry* 4:561-571.

Bell WE (1970). Management of temporomandibular joint problems. In: Current therapy in dentistry. GH M, GH W and RR Q editors. St Louis: CV Mosby Co, pp. 398-415.

Bindman AB, Keane D, Lurie N (1990). Measuring health changes among severely ill patients. The floor phenomenon. *Medical Care* 28:1142-1152.

Block SL (1980). Differential diagnosis of craniofacial-cervical pain. In: The temporomandibular joint. BG Sarnat and DM Laskin editors. Springfield: Charles C. Thomas, pp. 348-421.

Bolger N, Davis A, Rafaeli E (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology* 54:579-616.

Bower GH (1981). Mood and memory. *The American psychologist* 36:129-148.

Bowling A (2001). Measuring disease : a review of disease specific quality of life measurement scales. 2nd ed ed. Buckingham: Open University Press.

Bowling A (2002). Research methods in health: investigating health and health services. 2nd ed. Philadelphia, PA: Open University Press.

Bowling A (2005). Measuring health: a review of quality of life measurement scales. 3rd ed. Maidenhead: Open University Press.

Bowling A, Ebrahim S (2005). Handbook of health research methods: investigation, measurement and analysis. Maidenhead: Open University Press.

Boynton PM (2004). Hands-on guide to questionnaire research: Administering, analysing, and reporting your questionnaire. *British Medical Journal* 328:1372-1375.

Brennan DS, Spencer AJ (2006). Mapping oral health related quality of life to generic health state values. *BMC Health Services Research* 6.

Britten N (1995). Qualitative interviews in medical research. *British Medical Journal* 311:251-253.

Broderick JE, Schwartz JE, Shiffman S, Hufford MR, Stone AA (2003). Signaling Does Not Adequately Improve Diary Compliance. *Annals of Behavioral Medicine* 26:139-148.

Browne JP, O'Boyle CA, McGee HM, McDonald NJ, Joyce CR (1997). Development of a direct weighting procedure for quality of life domains. *Qual Life Res* 6:301-309.

Bryman A, Burgess RG (1994). Analyzing qualitative data London ; New York: Routledge.

Bullinger M, Anderson R, Cella D, Aaronson N (1993). Developing and evaluating cross-cultural instruments from minimum reqirements to optimal models. *Quality of Life Research* 2:451-459.

Bush FM, Harkins SW (1995). Pain-related limitation in activities of daily living in patients with chronic orofacial pain: psychometric properties of a disability index. *Journal of orofacial pain* 9:57-63.

Butterworth JC, Deardorff WW (1987). Psychometric profiles of craniomandibular pain patients: identifying specific subgroups. *Cranio : the journal of craniomandibular practice* 5:225-232.

Campbell DT, Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 56:81-105.

Clark GT, Delcanho RE, Goulet JP (1993). The utility and validity of current diagnostic procedures for defining temporomandibular disorder patients. *Advances in dental research* 7:97-112.

Cohen J (1988). Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates.

Cohen LK, Jago JD (1976). Toward the formulation of sociodental indicators. *International Journal of Health Services* 6:681-698.

Corson MA, Boyd T, Kind P, Allen PF, Steele JG (1999). Measuring oral health: does your treatment really make a difference. *Br Dent J* 187:481-484.

Coste J, Guillemin F, Pouchot J, Fermanian J (1997). Methodological approaches to shortening composite measurement scales. *J Clin Epidemiol* 50:247-252.

Costen JB (1934). A syndrome of ear and sinus symptoms dependent upon disturbed function of the temporomandibular joint. *Ann Otol Rhinol Laryngol* 43:1-15.

Couper MP (2005). Technology trends in survey data collection. *Social Science Computer Review* 23:486-501.

Crichton NJ (1993). The importance of statistics in research design. In: Clinical research methodology for complementary therapies. G Lewith and D Aldridge editors. London: Hodder & Stoughton.,

Crombez G, Vlaeyen JWS, Heuts PHTG, Lysens R (1999). Pain-related fear is more disabling than pain itself: Evidence on the role of pain-related fear in chronic back pain disability. *Pain* 80:329-339.

Cruise CE, Broderick J, Porter L, Kaell A, Stone AA (1996). Reactive effects of diary self-assessment in chronic pain patients. *Pain* 67:253-258.

Cunny KA, Perri M, 3rd (1991). Single-item vs multiple-item measures of health-related quality of life. *Psychol Rep* 69:127-130.

Cushing AM, Sheiham A, Maizels J (1986). Developing socio-dental indicators - the social impact of dental disease. *Community dental health* 3:3-17.

Dahlstrom L, Widmark G, Carlsson SG (1997). Cognitive-behavioral profiles among different categories of orofacial pain patients: Diagnostic and treatment implications. *European Journal of Oral Sciences* 105:377-383.

Dahlstrom L, Widmark G, Carlsson SG (2000). Changes in function and in pain-related and cognitive-behavioral variables after arthroscopy of temporomandibular joints. *European Journal of Oral Sciences* 108:14-21.

Dao TT, Lund JP, Lavigne GJ (1994). Comparison of pain and quality of life in bruxers and patients with myofascial pain of the masticatory muscles. *Journal of orofacial pain* 8:350-356.

Dawson J, Fitzpatrick R, Carr A, Murray D (1996). Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 78:185-190.

de Boer AG, van Lanschot JJ, Stalmeier PF, van Sandick JW, Hulscher JB, de Haes JC, *et al.* (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Qual Life Res* 13:311-320.

De Gagne TA, Mikail SF, D'Eon JL (1995). Confirmatory factor analysis of a 4-factor model of chronic pain evaluation. *Pain* 60:195-202.

de Oliveira CM, Sheiham A (2004). Orthodontic treatment and its impact on oral health-related quality of life in Brazilian adolescents. *Journal of orthodontics* 31:20-27; discussion 15.

De Vaus DA (2001). Surveys in social research. 5th ed. London: Routledge.

Derogatis LR, Cleary PA (1977). Confirmation of the dimensional structure of the SCL-90: a study in construct validation. *Journal of Clinical Psychology* 33:981-989.

Di Fabio RP (1998). Physical Therapy for Patients with TMD: A Descriptive Study of Treatment, Disability, and Health Status. *Journal of Orofacial Pain* 12:124-135.

Diatchenko L, Slade GD, Nackley AG, Bhalang K, Sigurdsson A, Belfer I, *et al.* (2005). Genetic basis for individual variations in pain perception and the development of a chronic pain condition. *Human Molecular Genetics* 14:135-143.

Dolan TA, Gooch BF, Bourque LB (1991). Associations of self-reported dental health and general health measures in the Rand Health Insurance Experiment. *Community Dent Oral Epidemiol* 19:1-8.

Dolan TA, Peek CW, Stuck AE, Beck JC (1998). Three-year changes in global oral health rating by elderly dentate adults. *Community Dentistry and Oral Epidemiology* 26:62-69.

Dworkin SF, Huggins KH, LeResche L, Von Korff M, Howard J, Truelove E, *et al.* (1990). Epidemiology of signs and symptoms in temporomandibular disorders: clinical signs in cases and controls. *The Journal of the American Dental Association* 120:273-281.

Dworkin SF, LeResche L (1992). Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. *Journal of craniomandibular disorders: facial & oral pain* 6:301-355.

Dworkin SF, Massoth DL (1994). Temporomandibular disorders and chronic pain: disease or illness? *The Journal of prosthetic dentistry* 72:29-38.

Dworkin SF, Sherman J, Mancl L, Ohrbach R, LeResche L, Truelove E (2002a). Reliability, validity, and clinical utility of the research diagnostic criteria for Temporomandibular Disorders Axis II Scales: depression, non-specific physical symptoms, and graded chronic pain. *J Orofac Pain* 16:207-220.

Dworkin SF, Turner JA, Mancl L, Wilson L, Massoth D, Muggins KH, *et al.* (2002b). A randomized clinical trial of a tailored comprehensive care treatment program for temporomandibular disorders. *Journal of Orofacial Pain* 16:259-276.

Eich E, Reeves JL, Jaeger B, Graff-Radford SB (1985). Memory for pain: Relation between past and present pain intensity. *Pain* 23:375-379.

Eversole LR, Machado L (1985). Temporomandibular joint internal derangements and associated neuromuscular disorders. *J Am Dent Assoc* 110:69-79.

Farrar JT (2000). What is clinically meaningful: outcome measures in pain clinical trials. *Clin J Pain* 16:S106-112.

Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL (2000). Defining the clinically important difference in pain outcome measures. *Pain* 88:287-294.

Farrar JT, Berlin JA, Strom BL (2003). Clinically important changes in acute pain outcome measures: a validation study. *J Pain Symptom Manage* 25:406-411.

Farrar WB (1972). Differentiation of temporomandibular joint dysfunction to simplify treatment. *J Prosthet Dent* 28:629-636.

Fayers PM, Hays RD (2004). Assessing quality of life in clinical trials: methods and practice. 2nd ed. Oxford: Oxford University Press.

Feine JS, Lund JP (1997). An assessment of the efficacy of physical therapy and physical modalities for the control of chronic musculoskeletal pain. *Pain* 71:5-23.

Fernandes MJ, Ruta DA, Ogden GR, Pitts NB, Ogston SA (2006). Assessing oral health-related quality of life in general dental practice in Scotland: Validation of the OHIP-14. *Community Dentistry and Oral Epidemiology* 34:53-62.

Fink A (1995). How to sample in surveys. In: The survey kit series. A Fink editor. Thousand Oaks, California: Sage Publications, pp. 73 v.

Fink A, Kosecoff JB (1998). How to conduct surveys: a step by step guide. 2nd ed. Thousand Oaks, CA: Sage Publications.

Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D (1992a). Quality of life measures in health care. I: Applications and issues in assessment. *Bmj* 305:1074-1077.

Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A (1992b). Importance of sensitivity to change as a criterion for selecting health status measures. *Qual Health Care* 1:89-93.

Fitzpatrick R, Boulton M (1994). Qualitative methods for assessing health care. *Quality in health care: QHC* 3:107-113.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess* 2:i-iv, 1-74.

Fitzpatrick R (1999). Assessment of quality of life as an outcome: finding measurements that reflect individuals' priorities. *Qual Health Care* 8:1-2.

Forouzanfar T, Weber WE, Kemler M, van Kleef M (2003). What is a meaningful pain reduction in patients with complex regional pain syndrome type 1? *Clin J Pain* 19:281-285.

Forssell H, Kirveskari P, Kangasniemi P (1985). Changes in headache after treatment of mandibular dysfunction. *Cephalalgia* 5:229-236.

Forssell H, Kalso E, Koskela P, Vehmanen R, Puukka P, Alanen P (1999). Occlusal treatments in temporomandibular disorders: a qualitative systematic review of randomized controlled trials. *Pain* 83:549-560.

Forssell H, Santalahti P, Puukka P, Talo S (2005). Searching for an assessment instrument to determine temporomandibular disorder pain profiles for the purposes of primary health care. *International Journal of Rehabilitation Research* 28:203-209.

Franklin B, Osborne H (1971). Research Methods: issues and insights. Belmont, CA: Wadsworth.

Freedman BA, Cohen SP, Kuklo TR, Lehman RA, Larkin P, Giuliani JR (2003). Intradiscal electrothermal therapy (IDET) for chronic low back pain in active-duty soldiers: 2-year follow-up. *Spine J* 3:502-509.

Fricton JR, Schiffman EL (1986). Reliability of a craniomandibular index. *J Dent Res* 65:1359-1364.

Gantley M, Carter Y, Shaw S, Thomas C (1999). An introduction to qualitative methods for health professionals London: Royal College of General Practitioners.

Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R (2002). Quality of life measurement: bibliographic study of patient assessed health outcome measures. *Bmj* 324:1417.

Geigle R, Jones SB (1990). Outcomes measurement: a report from the front. *Inquiry* 27:7-13.

Gendreau M, Hufford MR, Stone AA (2003). Measuring clinical pain in chronic widespread pain: selected methodological issues. *Best Pract Res Clin Rheumatol* 17:575-592.

Gift HC, Atchison KA (1995). Oral health, health, and health-related quality of life. *Medical Care* 33.

Gill TM, Feinstein AR (1994). A critical appraisal of the quality of quality-of-life measurements. *Journal of the American Medical Association* 272:619-626.

Goldsmith CH, Boers M, Bombardier C, Tugwell P (1993). Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. *J Rheumatol* 20:561-565.

Goossens ME, Rutten-van Molken MP, Vlaeyen JW, van der Linden SM (2000). The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol* 53:688-695.

Goulet JP, Lavigne GJ, Lund JP (1995). Jaw pain prevalence among French-speaking Canadians in Quebec and related symptoms of temporomandibular disorders. *Journal of Dental Research* 74:1738-1744.

Guyatt GH, Bombardier C, Tugwell PX (1986). Measuring disease-specific quality of life in clinical trials. *Canadian Medical Association Journal* 134:889-895.

Guyatt GH, Kirshner B, Jaeschke R (1992). Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 45:1341-1345.

Guyatt GH, Feeny DH, Patrick DL (1993). Measuring health-related quality of life. *Ann Intern Med* 118:622-629.

Hakeberg M, Hallberg LRM, Berggren U (2003). Burning mouth syndrome: Experiences from the perspective of female patients. *European Journal of Oral Sciences* 111:305-311.

Hardt J, Gerbershagen HU, Franke P (2000). The symptom check-list, SCL-90-R: Its use and characteristics in chronic pain patients. *European Journal of Pain* 4:137-148.

Harris RE, Williams DA, McLean SA, Sen A, Hufford M, Gendreau RM, *et al.* (2005). Characterization and consequences of pain variability in individuals with fibromyalgia. *Arthritis Rheum* 52:3670-3674.

Helkimo M (1974a). Studies on function and dysfunction of the masticatory system. II. Index for anamnestic and clinical dysfunction and occlusal state. *Sven Tandlak Tidskr* 67:101-121.

Helkimo M (1974b). Studies on function and dysfunction of the masticatory system. I. An epidemiological investigation of symptoms of dysfunction in Lapps in the north of Finland. *Proc Finn Dent Soc* 70:37-49.

Henson BS, Inglehart MR, Eisbruch A, Ship JA (2001). Preserved salivary output and xerostomia-related quality of life in head and neck cancer patients receiving parotid-sparing radiotherapy. *Oral Oncology* 37:84-93.

Heydecke G, Penrod JR, Takanashi Y, Lund P, Feine JS, Thomason JM (2005). Cost-effectiveness of mandibular two-implant overdentures and conventional dentures in the edentulous elderly. *Journal of Dental Research* 84:794-799.

Hill S, Harries U, Popay J (1996). Is the short form 36 (SF-36) suitable for routine health outcomes assessment in health care for older people? Evidence from preliminary work in community based health services in England. *J Epidemiol Community Health* 50:94-98.

Holroyd KA, Malinoski P, Davis MK, Lipchik GL (1999). The three dimensions of headache impact: pain, disability and affective distress. *Pain* 83:571-578.

Jenkinson C (1994). Measuring health and medical outcomes. In: Social research today series ; v. 3. London ; Bristol, Pa.: UCL Press, pp. viii, 215.

Johansson A, Unell L, Carlsson GE, Söderfeldt B, Halling A (2003). Gender difference in symptoms related to temporomandibular disorders in a population of 50-year-old subjects. *Journal of Orofacial Pain* 17:29-35.

John MT, Patrick DL, Slade GD (2002). The German version of the Oral Health Impact Profile - translation and psychometric properties. *Eur J Oral Sci* 110:425-433.

John MT, LeResche L, Koepsell TD, Hujoel P, Miglioretti DL, Micheelis W (2003). Oral health-related quality of life in Germany. *Eur J Oral Sci* 111:483-491.

John MT, Hujoel P, Miglioretti DL, Leresche L, Koepsell TD, Micheelis W (2004a). Dimensions of oral-health-related quality of life. *J Dent Res* 83:956-960.

John MT, Micheelis W, Biffar R (2004b). Reference values in oral health-related quality of life for the abbreviated version of the Oral Health Impact Profile
[Normwerte mundgesundheitsbezogener Lebensqualität für Kurzversionen des Oral Health Impact Profile.]. *Schweizer Monatsschrift fur Zahnmedizin = Revue mensuelle suisse d'odonto-stomatologie = Rivista mensile svizzera di odontologia e stomatologia / SSO* 114:784-791.

John MT, Hirsch C, Reiber T, Dworkin SF (2006a). Translating the Research Diagnostic Criteria for Temporomandibular Disorders into German: Evaluation of content and process. *Journal of Orofacial Pain* 20:43-52.

John MT, Miglioretti DL, LeResche L, Koepsell TD, Hujoel P, Micheelis W (2006b). German short forms of the Oral Health Impact Profile. *Community Dentistry and Oral Epidemiology* 34:277-288.

John MT, Reißmann DR, Schierz O, Wassell RW (2007). Oral health-related quality of life in patients with temporomandibular disorders. *Journal of Orofacial Pain* 21:46-54.

Johnson JR, Temple R (1985). Food and Drug Administration requirements for approval of new anticancer drugs. *Cancer Treat Rep* 69:1155-1159.

Jokovic A, Locker D, Stephens M, Kenny D, Tompson B, Guyatt G (2002). Validity and reliability of a questionnaire for measuring child oral-health-related quality of life. *Journal of Dental Research* 81:459-463.

Jones JA, Orner MB, Spiro A, 3rd, Kressin NR (2003). Tooth loss and dentures: patients' perspectives. *Int Dent J* 53:327-334.

Juniper EF, Guyatt GH, Jaeschke R (1996). How to develop and validate a new health-related quality of life instrument. In: Quality of life and pharmacoeconomics in clinical trials. B Spilker editor. Philadelphia: Lippincott-Raven Publishers.

Juniper EF, Guyatt GH, Streiner DL, King DR (1997). Clinical impact versus factor analysis for quality of life questionnaire construction. *Journal of Clinical Epidemiology* 50:233-238.

Kelly M, Steele, J., Nuttall, N., Bradnock, G., Morris, J., Nunn, J. (2000). Adult Dental Health Survey: Oral Health in the United Kingdom 1998. London: The Stationary Office.

Kerns RD, Turk DC, Rudy TE (1985). The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain* 23:345-356.

Kirshner B, Guyatt G (1985). A methodological framework for assessing health indices. *Journal of Chronic Diseases* 38:27-36.

Kohlmann T (2002). Epidemiology of orofacial pain. *Epidemiologie orofazialer schmerzen* 16:339-345.

Kressin N, Spiro A, 3rd, Bosse R, Garcia R, Kazis L (1996). Assessing oral health-related quality of life: findings from the normative aging study. *Med Care* 34:416-427.

Kropmans TJB, Dijkstra PU, Van Veen A, Stegenga B, De Bont LGM (1999). The smallest detectable difference of mandibular function impairment in patients with a painfully restricted temporomandibular joint. *Journal of Dental Research* 78:1445-1449.

Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics* 33:159-174.

Larsson P, List T, Lundström I, Marcusson A, Ohrbach R (2004). Reliability and validity of a Swedish version of the Oral Health Impact Profile (OHIP-S). *Acta Odontologica Scandinavica* 62:147-152.

Leao A, Sheiham A (1996). The development of a socio-dental measure of dental impacts on daily living. *Community Dental Health* 13:22-26.

LeResche L (1997a). Assessment of physical and behavioral outcomes of treatment. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 83:82-86.

LeResche L (1997b). Epidemiology of temporomandibular disorders: Implications for the investigation of etiologic factors. *Critical Reviews in Oral Biology and Medicine* 8:291-305.

LeResche L, Mancl L, Sherman JJ, Gandara B, Dworkin SF (2003). Changes in temporomandibular pain and other symptoms across the menstrual cycle. *Pain* 106:253-261.

Levitt SR, McKinney MW (1994). Validating the TMJ scale in a national sample of 10,000 patients: demographic and epidemiologic characteristics. *Journal of orofacial pain* 8:25-35.

List T, Helkimo M (1995). A scale for measuring the activities of daily living (ADL) of patients with craniomandibular disorders. *Swedish dental journal* 19:33-40.

Litwin M (1995). How to measure survey reliability and validity. In: The survey kit series. A Fink editor. Thousand Oaks, California: Sage Publications, pp. 87 v.

Locker D (1988). Measuring oral health: a conceptual framework. *Community dental health* 5:3-18.

Locker D, Slade G (1993). Oral health and the quality of life among older adults: the oral health impact profile. *Journal (Canadian Dental Association)* 59:830-833, 837.

Locker D, Miller Y (1994). Evaluation of subjective oral health status indicators. *Journal of public health dentistry* 54:167-176.

Locker D, Slade G (1994). Association between clinical and subjective indicators of oral health status in an older adult population. *Gerodontology* 11:108-114.

Locker D, Jokovic A (1996). Using subjective oral health status indicators to screen for dental care needs in older adults. *Community Dent Oral Epidemiol* 24:398-402.

Locker D (1998). Issues in measuring change in self-perceived oral health status. *Community Dentistry and Oral Epidemiology* 26:41-47.

Locker D, Matear D, Stephens M, Lawrence H, Fayne B (2001). Comparison of the GOHAI and OHIP-14 as measures of the oral health-related quality of life of the elderly. *Community Dentistry and Oral Epidemiology* 29:373-381.

Locker D, Allen PF (2002). Developing short-form measures of oral health-related quality of life. *J Public Health Dent* 62:13-20.

Locker D, Gibson B (2006). The concept of positive health: A review and commentary on its application in oral health research. *Community Dentistry and Oral Epidemiology* 34:161-173.

Lund JP, Widmer CG, Feine JS (1995). Validity of diagnostic and monitoring tests used for temporomandibular disorders. *J Dent Res* 74:1133-1143.

Lundeen TF, Levitt SR, McKinney MW (1986). Discriminative ability of the TMJ scale: age and gender differences. *The Journal of prosthetic dentistry* 56:84-92.

Mason L, Moore RA, Edwards JE, Derry S, McQuay HJ (2004a). Topical NSAIDs for acute pain: a meta-analysis. *BMC Fam Pract* 5:10.

Mason L, Moore RA, Edwards JE, Derry S, McQuay HJ (2004b). Topical NSAIDs for chronic musculoskeletal pain: systematic review and meta-analysis. *BMC Musculoskelet Disord* 5:28.

Maxwell C (1978). Sensitivity and accuracy of the visual analogue scale: a psycho-physical classroom experiment. *Br J Clin Pharmacol* 6:15-24.

Mays N, Pope C (1995). Rigour and qualitative research. *British Medical Journal* 311:109-112.

McColl E, Thomas R, Royal College of General Practitioners. (2000). The use and design of questionnaires. London: Royal College of General Practioners.

McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, *et al.* (2001). Design and use of questionnaires : a review of best practice applicable to surveys of health service staff and patients Alton: Core Research on behalf of the NCCHTA.

McCracken G, Janssen J, Heasman L, Stacey F, Steen N, deJager M, *et al.* (2005). Assessing adherence with toothbrushing instructions using a data logger toothbrush. *Br Dent J* 198:29-32; discussion 24.

McGrath C, Bedi R (2001). An evaluation of a new measure of oral health related quality of life - OHQoL-UK(W). *Community Dental Health* 18:138-143.

McGrath C, Bedi R (2002). Population based norming of the UK oral health related quality of life measure (OHQoL-UK). *Br Dent J* 193:521-524; discussion 517.

McGrath C, Bedi R (2003a). Measuring the impact of oral health on quality of life in Britain using OHQoL-UK(W). *J Public Health Dent* 63:73-77.

McGrath C, Bedi R (2003b). Dental services and perceived oral health: are patients better off going private? *J Dent* 31:217-221.

McGrath C, Comfort MB, Lo ECM, Luo Y (2003). Patient-centred outcome measures in oral surgery: Validity and sensitivity. *British Journal of Oral and Maxillofacial Surgery* 41:43-47.

McGrath C, Bedi R (2004). Why are we "weighting"? An assessment of a self-weighting approach to measuring oral health-related quality of life. *Community Dent Oral Epidemiol* 32:19-24.

McGrath C, Wong AHH, Lo ECM, Cheung CS (2005). The sensitivity and responsiveness of an oral health related quality of life measure to tooth whitening. *Journal of Dentistry* 33:697-702.

McNeill C, Danzig WM, Farrar WB, Gelb H, Lerman MD, Moffett BC, *et al.* (1980). Position paper of the American Academy of Craniomandibular Disorders. Craniomandibular (TMJ) disorders - the state of the art. *The Journal of prosthetic dentistry* 44:434-437.

McNeill C (1983). Craniomandibular (TMJ) disorders - the state of the art. Part II: accepted diagnostic and treatment modalities. *The Journal of prosthetic dentistry* 49:393-397.

McNeill C, American Academy of Orofacial Pain, American Academy of Craniomandibular Disorders (1993). Temporomandibular disorders : guidelines for classification, assessment, and management.141.

McQuay HJ (2005). Systematic reviews. In: Pain 2005 - an updated review: refresher course syllabus. Seattle: IASP Press, pp. 259-264.

Meenan RF, Gertman PM, Mason JH (1980). Measuring health status in arthritis. The arthritis impact measurement scales. *Arthritis Rheum* 23:146-152.

Melzack R (1975). The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1:277-299.

Melzack R (1987). The short-form McGill pain questionnaire. *Pain* 30:191-197.

Menghoel AM, Heggen K (2004). Recovery from fibromyalgia - Previous patients' own experiences. *Disability and Rehabilitation* 26:46-53.

Merskey H, Bogduk N, International Association for the Study of Pain. Task Force on Taxonomy (1994). Classification of chronic pain : descriptions of chronic pain syndromes and definitions of pain terms. 2nd ed. Seattle: IASP Press.

Mohl ND, Lund JP, Widmer CG, McCall Jr. WD (1990). Devices for the diagnosis and treatment of temporomandibular disorders. Part II: Electromyography and sonography. *The Journal of prosthetic dentistry* 63:332-336.

Moore A, Moore O, McQuay H, Gavaghan D (1997). Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics: use of pain intensity and visual analogue scales. *Pain* 69:311-315.

Moran LA, Guyatt GH, Norman GR (2001). Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *Journal of Clinical Epidemiology* 54:571-579.

Mumcu G, Inanc N, Ergun T, Ikiz K, Gunes M, Islek U, *et al.* (2006). Oral health related quality of life is affected by disease activity in Behcet's disease. *Oral Diseases* 12:145-151.

Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P (1998). Qualitative research methods in health technology assessment: A review of the literature. *Health Technology Assessment* 2.

Murray H, Locker D, Mock D, Tenenbaum HC (1996). Pain and the quality of life in patients referred to a craniofacial pain unit. *J Orofac Pain* 10:316-323.

National Institute for Health and Clinical Excellence (2006). Methods for development of NICE public health guidance. http://www.nice.org.uk/ (last accessed 2006).

Newton-John T (2002). Measurment of pain in adults. In: Assessment and management of orofacial pain. JM Zakreweska and SD Harrison editors. Amsterdam: Elsevier Science B.v.

Ng SKS, Leung WK (2006). Oral health-related quality of life and periodontal status. *Community Dentistry and Oral Epidemiology* 34:114-122.

Nord E (1992). Methods for quality adjustment of life years. *Social Science and Medicine* 34:559-569.

Nunnally J, Berbstein I (1994). Psychometric Theory. 3rd ed ed. New York: McGraw-Hill.

Nuttall NM, Steele JG, Pine CM, White D, Pitts NB (2001). The impact of oral health on people in the UK in 1998. *Br Dent J* 190:121-126.

Okeson JP (1997). Current terminology and diagnostic classification schemes. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics* 83:61-64.

Okeson JP (2003). Management of temporomandibular disorders and occlusion. 5th ed. St. Louis, Mo.: Mosby.

Olsen S, Nolan MF, Kori S (1992). Pain measurement. An overview of two commonly used methods. *Anesthesiol Rev* 19:11-15.

Osaba D (2005). Health-related quality of life outcomes in clinical trials. In: Assessing quality of life in clinical trials. P Fayers and R Hays editors. Oxford: Oxford University Press.

Pace-Balzan A, Cawood JI, Howell R, Butterworth CJ, Lowe D, Rogers SN (2006). The further development and validation of the Liverpool Oral Rehabilitation Questionnaire: a cross-sectional survey of patients attending for oral rehabilitation and general dental practice. *Int J Oral Maxillofac Surg* 35:72-78.

Pallant J (2005). SPSS survival manual : a step by step guide to data analysis using SPSS for Windows (version 12). 2nd ed. Maidenhead: Open University Press.

Patrick DL, Deyo RA (1989). Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 27:S217-232.

Pau AKH, Croucher R, Marcenes W (2000). Perceived inability to cope and care-seeking in patients with toothache: A qualitative study. *British Dental Journal* 189:503-506.

Pehling J, Schiffman E, Look J, Shaefer J, Lenton P, Fricton J (2002). Interexaminer reliability and clinical validity of the temporomandibular index: a new outcome measure for temporomandibular disorders. *J Orofac Pain* 16:296-304.

Petrie A, Bulman JS, Osborn JF (2002). Further statistics in dentistry. Part 4: Clinical trials 2. *British Dental Journal* 193:557-561.

Pollard CA (1984). Preliminary validity study of the pain disability index. *Perceptual and Motor Skills* 59:974.

Porro CA (2003). Functional imaging and pain: Behavior, perception, and modulation. *Neuroscientist* 9:354-369.

Pow EHN, Leung KCM, McMillan AS (2001). Prevalence of Symptoms Associated with Temporomandibular Disorders in Hong Kong Chinese. *Journal of Orofacial Pain* 15:228-234.

Price DD, McGrath PA, Rafii A, Buckingham B (1983). The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain* 17:45-56.

RAND (2006). The 36-item Short Form Health Survey Instrument (SF-36). http://www.rand.org/health/surveys/sf36item/question.html (last accessed 2006).

Raymond MC, Brown JB (2000). Experience of fibromyalgia: Qualitative study. *Canadian Family Physician* 46:1100-1106.

Redelmeier DA, Kahneman D (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 66:3-8.

Reisine ST (1984). Dental disease and work loss. *Journal of Dental Research* 63:1158-1161.

Reisine ST (1985). Dental health and public policy: The social impact of dental disease. *American Journal of Public Health* 75:27-30.

Reisine ST, Fertig J, Weber J, Leder S (1989). Impact of dental conditions on patients' quality of life. *Community dentistry and oral epidemiology* 17:7-10.

Reisine ST, Weber J (1989). The effects of temporomandibular joint disorders on patients' quality of life. *Community dental health* 6:257-270.

Riffenburgh R (2006). Statistics in medicine. Massachusetts: Elsevier.

Robinson PG, Gibson B, Khan FA, Birnbaum W (2001). A comparison of OHIP 14 and OIDP as interviews and questionnaires. *Community Dental Health* 18:144-149.

Rogers SN, Humphris G, Lowe D, Brown JS, Vaughan ED (1998). The impact of surgery for oral cancer on quality of life as measured by the Medical Outcomes Short Form 36. *Oral Oncology* 34:171-179.

Rosenstiel AK, Keefe FJ (1983). The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. *Pain* 17:33-44.

Ross Beirne O (2004). Removal of mandibular third molars improved patients' quality of life. *Journal of Evidence-Based Dental Practice* 4:212-213.

Rosser RM, Watts VC (1972). The measurement of hospital output. *International Journal of Epidemiology* 1:361-368.

Rudy TE, Turk DC, Zaki HS, Curtin HD (1989). An empirical taxometric alternative to traditional classification of temporomandibular disorders. *Pain* 36:311-320.

Rudy TE, Turk DC, Kubinski JA, Zaki HS (1995). Differential treatment responses of TMD patients as a function of psychological characteristics. *Pain* 61:103-112.

Rudy TE, Greco CM, Yap GA, Zaki HS, Leader JK, Boston JR (2001). The association between research diagnostic criteria for temporomandibular disorder findings and biting force and endurance in patients with temporomandibular disorders. *Pain Medicine* 2:35-45.

Rugh JD, Solberg WK (1985). Oral health status in the United States: temporomandibular disorders. *Journal of dental education* 49:398-406.

Sato H (1997). ICIDH, New ICIDH and Asian / Pacific Decade. http://www.dinf.ne.jp/doc/english/asia/resource/z00ap/001/z00ap00109.htm (last accessed 2006).

Schiffman EL, Fricton JR, Haley DP, Shapiro BL (1990). The prevalence and treatment needs of subjects with temporomandibular disorders. *The Journal of the American Dental Association* 120:295-303.

Schweinhardt P, Lee M, Tracey I (2006). Imaging pain in patients: Is it meaningful? *Current Opinion in Neurology* 19:392-400.

Segù M, Collesano V, Lobbia S, Rezzani C (2005). Cross-cultural validation of a short form of the Oral Health Impact Profile for temporomandibular disorders. *Community Dentistry and Oral Epidemiology* 33:125-130.

Siddall PJ, Cousins MJ (2004). Persistent pain as a disease entity: Implications for clinical management. *Anesthesia and Analgesia* 99:510-520.

Slade GD, Spencer AJ (1994). Development and evaluation of the Oral Health Impact Profile. *Community Dent Health* 11:3-11.

Slade GD, Spencer AJ, Locker D, Hunt RJ, Strauss RP, Beck JD (1996). Variations in the social impact of oral conditions among older adults in South Australia, Ontario, and North Carolina. *Journal of Dental Research* 75:1439-1450.

Slade GD (1997a). Measuring Oral Health and Quality of Life. First International Conferencen on Measuring Oral Health., Univeristy of North Carolina-Chapel Hill: Department of Dental Ecology.

Slade GD (1997b). Derivation and validation of a short-form oral health impact profile. *Community Dent Oral Epidemiol* 25:284-290.

Slade GD (1998). Assessing change in quality of life using the Oral Health Impact Profile. *Community Dentistry and Oral Epidemiology* 26:52-61.

Slade GD (2002). Assessment of oral health-related quality of life. In: Oral health-related quality of life. M Inglehart and R Bagramian editors. Illinois: Quintessence Publishing.

Sprangers MA, Aaronson NK (1992). The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J Clin Epidemiol* 45:743-760.

Steele JG, Sanders AE, Slade GD, Allen PF, Lahti S, Nuttall N, *et al.* (2004). How do age and tooth loss affect oral health impacts and quality of life? A study comparing two national samples. *Community Dent Oral Epidemiol* 32:107-114.

Steele JG, Moufti MA, Wassell RW, Durham J, Meechan JG (2006). Does TMD Affect Quality of Life in Unexpected Ways? The Pan European Federation Meeting of the International Association of Dental Research. Trinity College Dublin, Ireland.

Stegenga B, de Bont LG, de Leeuw R, Boering G (1993). Assessment of mandibular function impairment associated with temporomandibular joint osteoarthrosis and internal derangement. *Journal of orofacial pain* 7:183-195.

Stone AA, Shiffman S (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine* 16:199-202.

Stone AA, Broderick JE, Kaell AT, Delespaul PAEG, Porter LE (2000). Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics? *Journal of Pain* 1:212-217.

Stone AA, Shiffman S (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine* 24:236-243.

Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR (2002). Patient non-compliance with paper diaries. *British Medical Journal* 324:1193-1194.

Strassburger C, Heydecke G, Kerschbaum T (2004). Influence of prosthetic and implant therapy on satisfaction and quality of life: a systematic literature review. Part 1- Characteristics of the studies. *Int J Prosthodont* 17:83-93.

Strauss RP, Hunt RJ (1993). Understanding the value of teeth to older adults: influences on the quality of life. *The Journal of the American Dental Association* 124:105-110.

Streiner DL, Norman GR (1995). Health measurement scales : a practical guide to their development and use. 2nd ed. Oxford: Oxford University Press.

Sugisaki M, Kino K, Yoshida N, Ishikawa T, Amagasa T, Haketa T (2005). Development of a new questionnaire to assess pain-related limitations of daily functions in Japanese patients with temporomandibular disorders. *Community Dentistry and Oral Epidemiology* 33:384-395.

Suvinen TI, Odont L, Hanes KR, Gerschman JA, Reade PC (1997a). Psychophysical subtypes of temporomandibular disorders. *Journal of Orofacial Pain* 11:200-205.

Suvinen TI, Reade PC, Sunden B, Gerschman JA, Koukounas E (1997b). Temporomandibular disorders: Part II. A comparison of psychologic profiles in Australian and Finnish patients. *Journal of Orofacial Pain* 11:147-157.

Suvinen TI, Reade PC, Sunden B, Gerschman JA, Koukounas E (1997c). Temporomandibular disorders. Part I: A comparison of symptom profiles in Australian and Finnish patients. *Journal of Orofacial Pain* 11:58-66.

Suvinen TI, Reade PC, Kemppainen P, Kononen M, Dworkin SF (2005). Review of aetiological concepts of temporomandibular pain disorders: towards a biopsychosocial model for integration of physical disorder factors with psychological and psychosocial illness impact factors. *Eur J Pain* 9:613-633.

Ta LE, Dionne RA (2004). Treatment of painful temporomandibular joints with a cyclooxygenase-2 inhibitor: A randomized placebo-controlled comparison of celecoxib to naproxen. *Pain* 111:13-21.

Talley RL, Murphy GJ, Smith SD, Baylin MA, Haden JL (1990). Standards for the history, examination, diagnosis, and treatment of temporomandibular disorders (TMD): A position paper. *Cranio - Journal of Craniomandibular Practice* 8:60-77.

Tarlov AR, Ware Jr JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M (1989). The Medical Outcomes Study. An application of methods for monitoring the results of medical care. *Journal of the American Medical Association* 262:925-930.

The Information Commissioner's Office in the UK (1998). Data protection guide. http://www.ico.gov.uk/Home/for_organisations/data_protection_guide.aspx (last accessed 2003).

Todd KH, Funk JP (1996). The minimum clinically important difference in physician-assigned visual analog pain scores. *Acad Emerg Med* 3:142-146.

Truelove EL, Sommers EE, LeResche L, Dworkin SF, Von Korff M (1992). Clinical diagnostic criteria for TMD. New classification permits multiple diagnoses. *J Am Dent Assoc* 123:47-54.

Turk DC (1997). Psychosocial and behavioral assessment of patients with temperomandibular disorders: Diagnostic and treatment implications. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics* 83:65-71.

Turk DC, Melzack R (2001). Pain measurment. In: Handbook of pain assessment. New York: Guilford Press, pp. xix, 760.

Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, *et al.* (2003). Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain* 106:337-345.

Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, *et al.* (2006). Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. *Pain* 125:208-215.

Van Wijk AJ, Hoogstraten J (2002). The McGill Pain Questionnaire in dentistry: Dental pain described in numbers. *Pain Reviews* 9:87-97.

Verbrugge LM (1980). Health diaries. *Medical Care* 18:73-95.

Von Korff M, Dworkin SF, Le Resche L (1990). Graded chronic pain status: An epidemiologic evaluation. *Pain* 40:279-291.

Ware JE, Jr., Sherbourne CD (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30:473-483.

Wassell RW, Adams N, Kelly PJ (2004). Treatment of temporomandibular disorders by stabilising splints in general dental practice: results after initial treatment. *Br Dent J* 197:35-41; discussion 31; quiz 50-31.

Wewers ME, Lowe NK (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health* 13:227-236.

Wig AD, Aaron LA, Turner JA, Huggins KH, Truelove E (2004). Short-term clinical outcomes and patient compliance with temporomandibular disorder treatment recommendations. *Journal of Orofacial Pain* 18:203-213.

Williams A (1990). EuroQol - A new facility for the measurement of health-related quality of life. *Health Policy* 16:199-208.

Wolfart S, Heydecke G, Luthardt RG, Marré B, Freesmeyer WB, Stark H, *et al.* (2005). Effects of prosthetic treatment for shortened dental arches on oral health-related quality of life, self-reports of pain and jaw disability: Results from the pilot-phase of a randomized multicentre trial. *Journal of Oral Rehabilitation* 32:815-822.

World Health Organization (1946). WHO definition of Health. http://www.who.int/about/definition/en/ (last accessed 2003).

World Health Organization (1980). International classification of impairments, disabilities and handicaps. Geneva.

World Health Organization (2006). The WHO Family of International Classifications. http://www.who.int/classifications/en/ (last accessed 2006).

Worthington HV (1998). Statistical aspects of measuring change in oral health status of older adults. *Community Dentistry and Oral Epidemiology* 26:48-51.

Yap AJ (2002). TMD Around the World: Initial Findings from the International Consortium for RDC/TMD - based Research. IADR/AADR/CADR 80th General Session (March 6-9, 2002) San Diego, California, Wednesday, 6 March 2002.

Yap AUJ, Tan KBC, Hoe JKE, Yap RHC, Jaffar J (2001). On-line computerized diagnosis of pain-related disability and psychological status of TMD patients: A pilot study. *Journal of Oral Rehabilitation* 28:78-87.

Yap AUJ, Tan KBC, Chua EK, Tan HH (2002). Depression and somatization in patients with temporomandibular disorders. *Journal of Prosthetic Dentistry* 88:479-484.

Yap AUJ, Chua EK, Tan KBC, Chan YH (2004). Relationships between depression/somatization and self-reports of pain and disability. *Journal of Orofacial Pain* 18:220-225.

Zealley AK, Aitken RC (1969). Measurement of mood. *Proc R Soc Med* 62:993-996.

Appendices

Appendix 1
TMD diagnostic criteria and examination forms

Appendix 1 is divided into four parts. It starts with a summary of the Research Diagnostic Criteria (RDC). These are presented in the original paper (Dworkin and LeResche, 1992) in a theoretical manner and over numerous pages. Therefore, it was necessary to summarise these in a one-sheet form to facilitate the operational procedures of making a diagnosis in busy dental and hospital clinics.

Presented next is the modified form of the Temporomandibular Index (TMI) (Pehling *et al.*, 2002). The advantages of this index are outlined in section 2.2.4.3. However, as discussed in that section, some work was needed to overcome the shortcomings of this index. The original form is complicated and contains procedures of dubious utility and reliability, for example, the differentiation between coarse and fine crepitus. In addition, it endeavours to produce a score for each patient, given equal weights to different signs.

The third part of this appendix is the Graded Chronic Pain Scale section of the History Questionnaire of the RDC.

Lastly, presented in part four are examples of the Arabic version of the RDC. The literature review of this thesis reported that one very important feature of the RDC is its worldwide appeal. RDC was available in 17 languages at the time of submitting this thesis. However, they were not all obtained using a scientific approach such as the *Forward - Backward Translation* method used for example to produce the German version (John *et al.*, 2006a). The Arabic version, with which the author of this thesis is familiar was clearly made by a non-native speaker and was poorly devised. Therefore, there is a need to produce a new version using recognised methods such as the above mentioned.

Research Diagnostic Criteria for the Temporomandibular Disorders RDC/TMD

Group	Criteria
Muscle Disorders	
Myofascial pain Key: Painful muscles	1. Reported pain in masticatory muscles [1] 2. Pain on palpation in at least 3 sites [2], one of them at least in the same side of the reported pain
Myofascial pain with limited opening Key: Painful muscles +limited movement	1. Myofascial pain 2. Pain-free unassisted [3] opening< 40 mm 3. Passive [4] stretch ≥ 5 mm (from pain-free unassisted opening to "painful" assisted opening)
Disc Displacements	
Disc displacement with reduction Key: Reproducible clicking	1. No pain in the joint neither reported nor on palpation 2. Reproducible [5] click on any excursion [6] with either opening or closing click 3. With click on opening and closing (unless excursive click confirmed): • Click on opening occurs at ≥ 5 mm interincisal distance than on closing • Clicks eliminated by protrusive opening
Disc displacement without reduction with limited opening Key: Limited opening with no clicking	1. History of locking or catching that interfered with eating 2. Absence of TMJ clicking meeting DDR criteria 3. Unassisted "painful" opening ≤ 35mm 4. Passive stretch <5 mm (from "painful" unassisted opening to "painful" assisted opening) 5. Contralateral excursion < 7mm ***Or*** Uncorrected ipsilateral deviation on opening
Disc displacement without reduction without limited opening Key: History of previously limited opening-imaging needed to confirm DD	1. History of locking or catching that interfered with eating 2. The presence of TMJ sounds excluding DDR clicking 3. Unassisted "painful" opening > 35mm 4. Passive stretch ≥ 5 mm (from "painful" unassisted opening to "painful" assisted opening) 5. Contralateral excursion ≥ 7mm 6. Optional imaging (Arthrography or MRI) to confirm DD
Other Common Joint Diseases [7]	
Arthralgia Key: Painful TMJ / no crepitus	1. Pain on TMJ palpation either laterally or intra auricular 2. Self reported joint pain with or without jaw movement 3. Absence of crepitus, and Possibility of clicking
Osteoarthritis Key: Painful TMJ + crepitus	1. Pain as for Arthralgia (reported and on palpation) 2. Crepitus on any movement ***or*** Tomogram evidence of joint changes [8]
Osteoarthrosis Key: Non painful TMJ + crepitus	1. Crepitus on any movement ***or*** Tomogram evidence of joint changes [8] 2. No reported joint pain , neither on palpation nor on any movement

1 In the jaw, temples, face, preauricular area, or inside the ear, at rest or function
2 There are 20 sites (ten on each side): posterior, middle and anterior Temporalis; origin, body and insertion of Masseter; posterior mandibular region; Submandibular region; lateral Pterygoid; tendon of Temporalis.
3 Interincisal opening plus overbite or interincisal opening minus anterior open bite.
4 Passive stretch: The examiner's index and thumb are used to moderately force the mouth to open wider than unassisted opening. Patients are instructed to raise a hand to signal when the stretch becomes too uncomfortable.
5 All clicks must be reproduced 2 out of 3 consecutive trials
6 Contra- or ipsilateral or protrusive
7 Erosion of cortical delineation, sclerosis of parts or all the condyle and articular eminence, flattening of joint surfaces, osteophyte formation
8 Erosion of cortical delineation, sclerosis of parts or all the condyle and articular eminence, flattening of joint surfaces, osteophyte formation

Patient's code _____ Date / / 20

TMD EXAMINATION

Examiner: Please follow the following instructions fill in all the form items

1. For all items please circle [0] if the exam item is negative and [1] if the exam item is positive
2. For Vertical incisal overlap use (+) for overbite, (-) for anterior open bite
3. Make sure there is no unstable denture

Facial Asymmetry [0] [1] *Specify if yes* _____

Muscle Hypertrophy [0] [1] *Specify if yes* _____

Do you have a pain on the right side of your face, the left side or both? **Right Left None Both**

Please circle one

Could you point to the areas where you feel pain?

Please circle one on each side

Right	**Left**
None	None
Jaw joint	Jaw joint
Muscles	Muscles
Both	Both

Final Diagnosis

Group/side	**Right**	**Left**
RDC I		
RDC II		
RDC III		

Joint examination

Palpate the following (1Kg pressure)

			Right	Left
Joint Pain	Tender lateral		[0] [1]	[0] [1]
	Tender ear canal		[0] [1]	[0] [1]
	Tender on any movement		[0] [1]	[0] [1]
Joint sounds	Opening click	Occurs at…….mm	[0] [1]	[0] [1]
	Closing click	Occurs at …….mm	[0] [1]	[0] [1]
	Reciprocal click eliminated by protrusive opening		[0] [1]	[0] [1]
	Excursion click (Contra – or ipsilateral)		[0] [1]	[0] [1]
	Protrusion click		[0] [1]	[0] [1]
	Crepitus		[0] [1]	[0] [1]

Muscle examination

Palpate the following (1Kg extra-orally, 0.5Kg intra-orally)

		Right	Left
Neck & Shoulder	Sternomastoid	[0] [1]	[0] [1]
	Posterior intrinsic neck	[0] [1]	[0] [1]
	Trapezius	[0] [1]	[0] [1]
	Shoulder	[0] [1]	[0] [1]
Extraoral muscles	Anterior Temporalis	[0] [1]	[0] [1]
	Middle Temporalis	[0] [1]	[0] [1]
	Posterior Temporalis	[0] [1]	[0] [1]
	Origin of the Masseter	[0] [1]	[0] [1]
	Body of the Masseter	[0] [1]	[0] [1]
	Insertion of the Masseter	[0] [1]	[0] [1]
	Posterior mandibular region	[0] [1]	[0] [1]
	Submandibular region	[0] [1]	[0] [1]
Intraoral muscles	Tendon of the Temporalis	[0] [1]	[0] [1]
	Lateral Pterygoid area	[0] [1]	[0] [1]

Total intra- and extra-oral muscles only ☐

Function examination

Examiner: please ask the subject to open 3 times

Deviation on opening [0] [1] To Right or Left? [R] [L] Uncorrected? [0] [1]

Please describe opening pattern: ……

Maxillary and Mandibular incisors used —+— Vertical incisal overlap (Please define + or -) ....mm
Opening movement including incisal overlap

Movement	**Amount**	**Painful movement**	**If painful, where is the pain located?**
Maximum pain-free unassisted opening	……..mm		
Maximum "painful" unassisted opening	……..mm	[0] [1]	……
Maximum "painful" assisted opening	……..mm	[0] [1]	……

Excursion movement (normal amount is ≥ 7 mm)

Movement	**Amount**	**Restricted excursion**	**Painful movement**	**If painful, where is the pain located?**
To right	.....mm	[0] [1]	[0] [1]	……
To left	……mm	[0] [1]	[0] [1]	……
Protrusion	……mm	[0] [1]	[0] [1]	……

Examiner: Please check steps 1 & 2 for accuracy

Graded Chronic Pain Scale

Please read each question and respond accordingly. For each of the questions below circle only one response.

1. How would you rate your facial pain on a 0 to 10 scale at the present time, that is right now, where 0 is "no pain" and 10 is "pain as bad as could be"?

No pain — Pain as bad as could be

0 1 2 3 4 5 6 7 8 9 10

2. In the past six months, how intense was your worst pain rated on a 0 to 10 scale where 0 is "no pain" and 10 is "pain as bad as could be"?

No pain — Pain as bad as could be

0 1 2 3 4 5 6 7 8 9 10

3. In the past six months, on the average, how intense was your pain rated on a 0 to 10 scale where 0 is "no pain" and 10 is "pain as bad as could be"? [That is, your usual pain at times you were experiencing pain].

No pain — Pain as bad as could be

0 1 2 3 4 5 6 7 8 9 10

4. In the past six months, how much has facial pain interfered with your daily activities rated on a 0 to 10 scale where 0 is "no interference" and 10 is "unable to carry on any activities"?

No interference — Unable to carry on any activities

0 1 2 3 4 5 6 7 8 9 10

5. In the past six months, how much has facial pain changed your ability to take part in recreational, social and family activities where 0 is "no change" and 10 is "extreme change"?

No change — Unable to carry on any activities

0 1 2 3 4 5 6 7 8 9 10

6. In the past six months, how much has facial pain changed your ability to work including housework) where 0 is "no change" and 10 is "extreme change"?

No change — Unable to carry on any activities

0 1 2 3 4 5 6 7 8 9 10

7. About how many days in the last six months have you been kept from your usual activities (work, school or housework) because of facial pain? ---------- Days

Examples from the Arabic version of the RDC

Presented here are examples of questions from the Arabic version of the RDC obtained from the RDC website (http://rdc-tmdinternational.org). This is followed by an English interpretation made by the author of this thesis.

يجب عليك الاشارة على الخط X في المكان الموازي بالنسبة لمقدار شدة الظاهرة. وضع ال- X في الطرف الشمالي للخط يشير الى انعدام الظاهرة ,بينما كلما اقتربت الاشارة الى الطرف اليميني للخط فازدادت شدة الظاهرة.

You have to mark on the line X in a place parallel in relation to the degree of the phenomenon's intensity. Putting the – X in the left side of the line indicates the absence of the phenomenon, while the closer the mark is to the right side of the line then the phenomenon is increasing.

8 . خلال الاشهر الستة الاخيرة ,ما هي شدة الوجع الأليم الذي كنت تعاني / تعانين منه ؟

|----------|

وجع غير محتمل ما من وجع

8. During the last 6 months, what is the intensity of the painfully ache that you have been suffering from? No pain --- Unbearable pain.

(This question corresponds to question No 2 in the Graded Chronic Pain Scale)

9 . خلال الاشهر الستة الاخيرة ,ما هي شدة الوجع المعدلة الذي كنت تعاني / تعانين منه ؟

|----------|

وجع غير محتمل ما من وجع

9. During the last 6 months, what is the intensity of the modified ache that you have been suffering from? No pain --- Unbearable pain.

(This question corresponds to question No 3 in the Graded Chronic Pain Scale)

11 . خلال الاشهر الستة الاخيرة ,لاي مقدار كان يزعج وجع وجهك القيام بنشاطاتك العادية ؟

|----------|

غير قادر على القيام باي نشاط بدون ازعاج

11. During the last 6 months, to what extent has been disturbing the pain of your face the performance to your usual activities? Have been unable to do any activity --- without annoyance

(This question corresponds to question No 4 in the Graded Chronic Pain Scale)

Appendix 2
Formula for calculating the area under the curve in *Study One*

Simon Kometa

Problem Description: I have a series of variables, named Y1 to Y4, in an SPSS data file. These variables represent outcomes at 4 ordered time points. For each case, I want to calculate the area under the curve of Y1 to Y4 as a function of time. The values of time are assumed to be 1, 2, 3, and 4, i.e. time is the sequence number of the Y variable. The result should be stored in a new variable in the active file. This area function will be equivalent to the AREA function that is available in in the BMDP command language. In BMDP, the /TRANSFORM paragraph command: yarea = area(y1,y2,y3,y4). calculates YAREA as (Y1 + 2*Y2 + 2*Y3 + Y4)/2. How can I implement such a function in SPSS?

Resolution Subject: SPSS transformation commands do not include a built-in area function such as the BMDP AREA function.

Resolution Description:
SPSS transformation commands do not include a built-in area function such as the BMDP AREA function. However, the desired area calculation can be easily performed with a combination of the functions available in SPSS. Assuming that the variables Y1 to Y4 are adjacent to each other in the open data file and that they are correctly ordered, the following command would store the area in the new variable YAREA.

```
IF (NMISS(y1 TO y4) = 0) yarea = (2*SUM(y1 to y4) - y1 - y4)/2 .
```

The command above can be implemented through the menus in SPSS versions with a graphic user interface.
1. From the Data Editor window, choose Transform->Compute from the menus.
2. Enter YAREA as the Target Variable.
3. Enter "(2*SUM(y1 to y4)-y1-y4)/2" into the Numeric Expression box (without the quotation marks).
4. Click the If button. In the dialog box that appears, click the radio button "Include if case satisfies condition".
5. Enter "NMISS(y1 TO y4) = 0" into the box underneath this choice (again, no quotation marks) and click Continue.
6. Click OK.

If the variables Y1 to Y4 are not adjacent and ordered in the data file, then you will need to list the all the variables as arguments for the NMISS and SUM functions, as in NMISS(y1,y2,y3,y4).

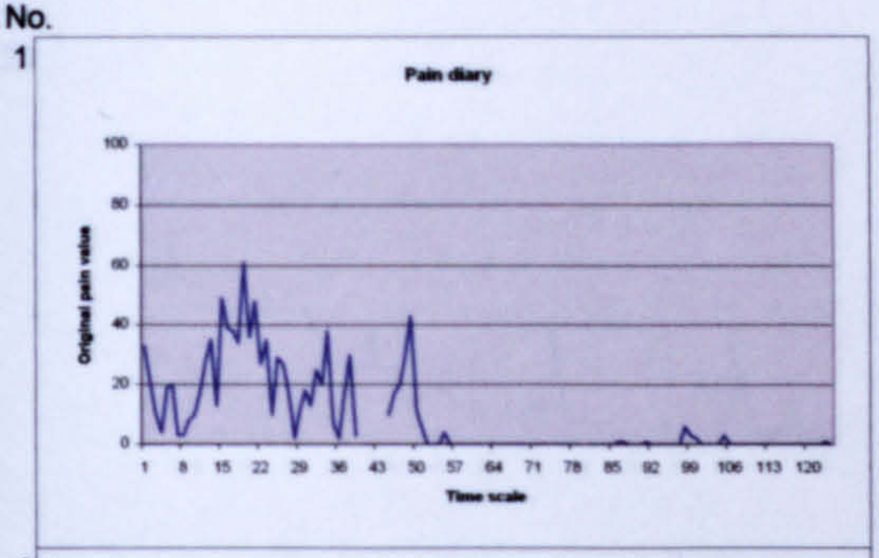
The IF condition checks that there are no missing values in the sequence for the current case. If there are missing values among the Y variables, then YAREA will be system missing for the current case. (This is also true for the AREA function in BMDP). It would be possible to write commands that would allow YAREA to be calculated in the presence of some missing values. This would involve a more extensive set of commands which would depend on the particular arrangements of missing values that the user considered allowable and the particular adjustment desired.

Appendix 3
Individual plots of pain/time response to treatment in *Study One*

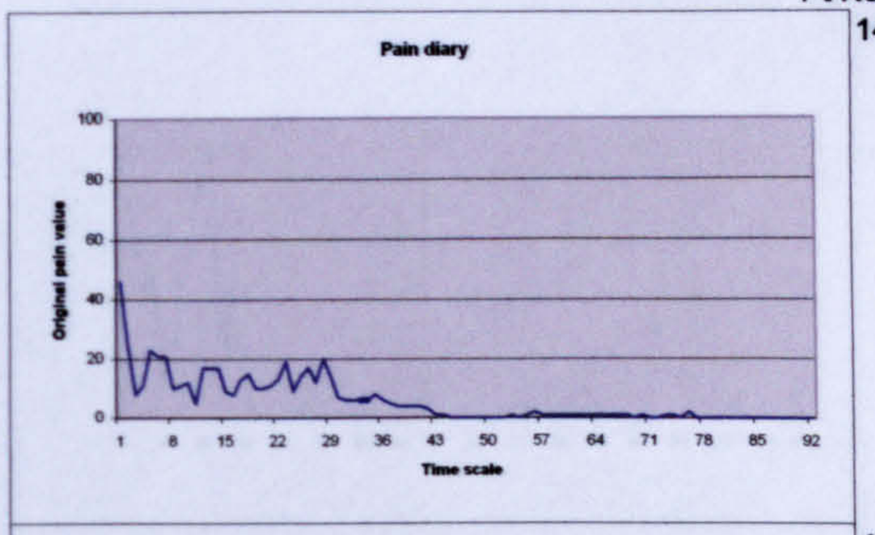
Improvers' charts

Page 1

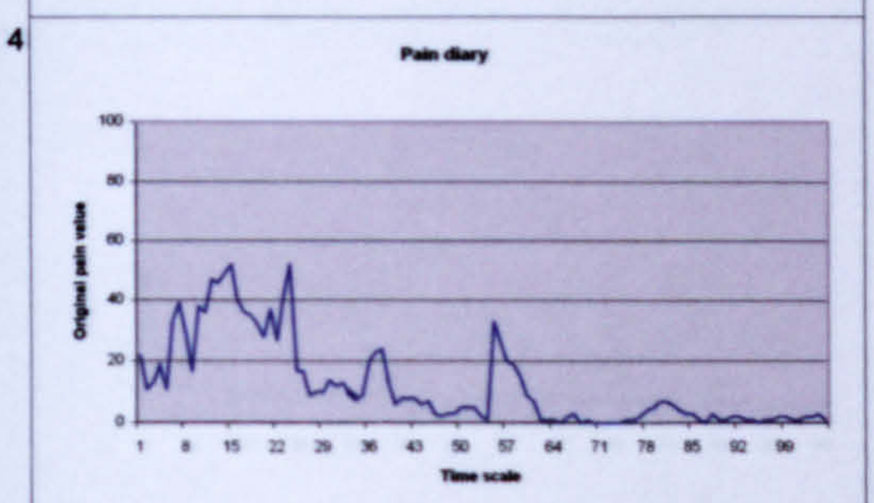
Pt No. 1



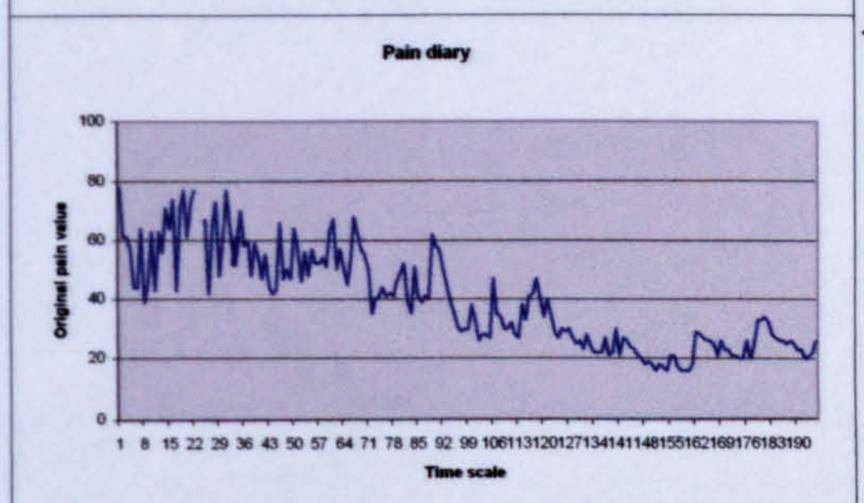
Pt No. 14



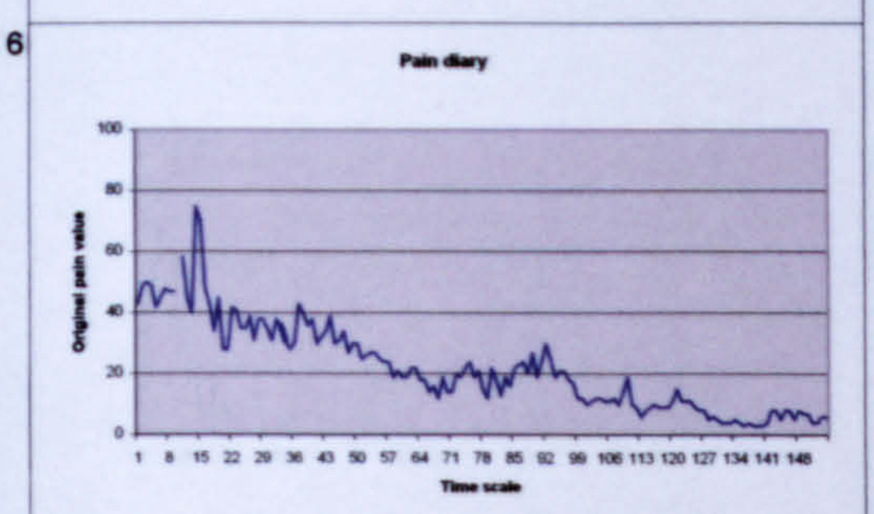
Pt No. 4



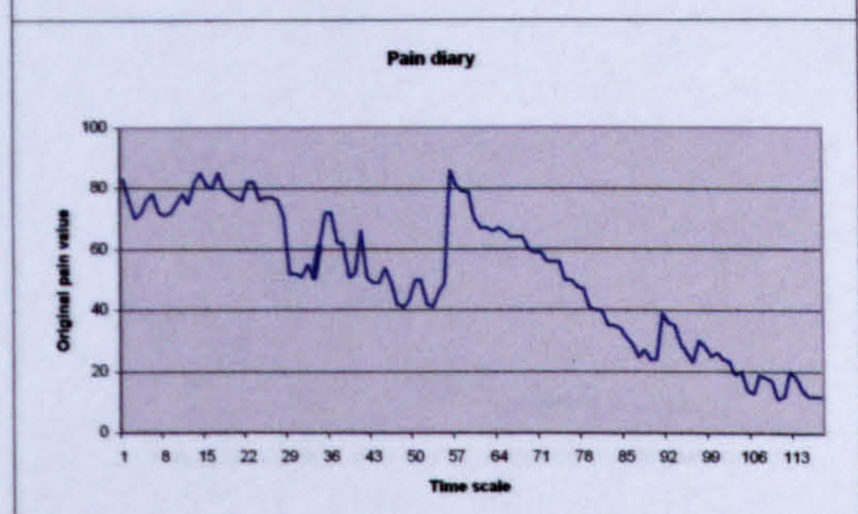
Pt No. 15



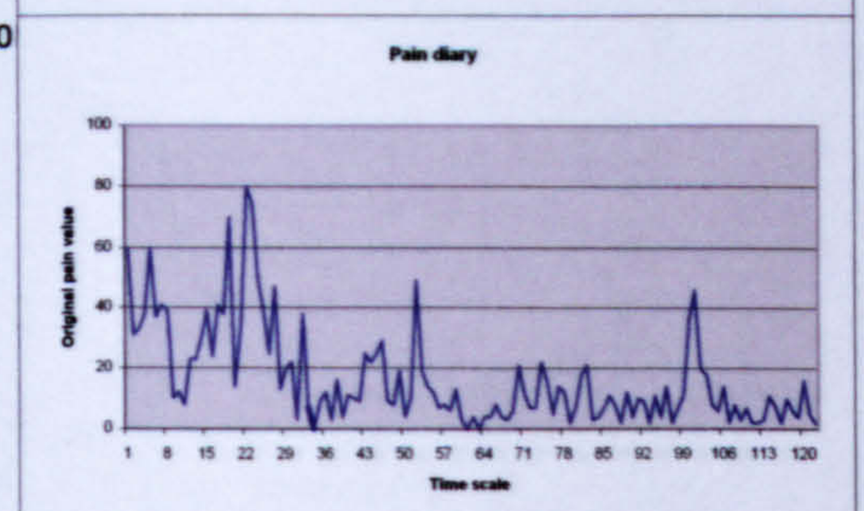
Pt No. 6



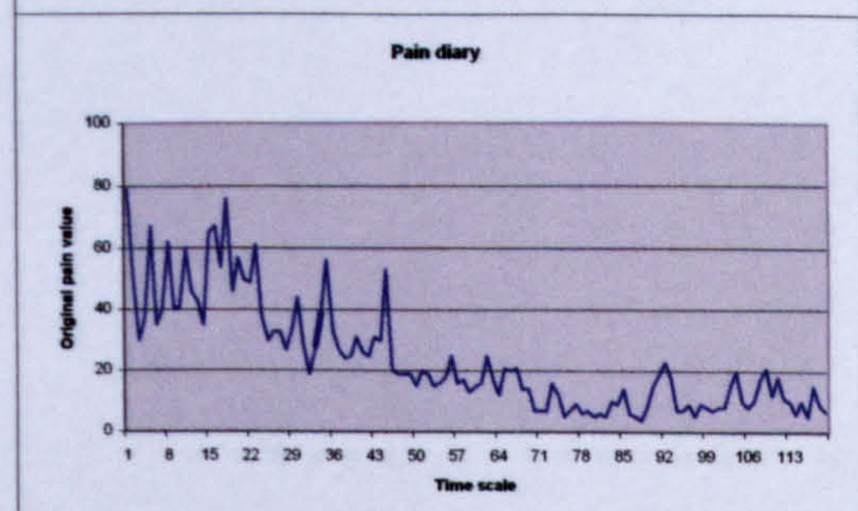
Pt No. 16



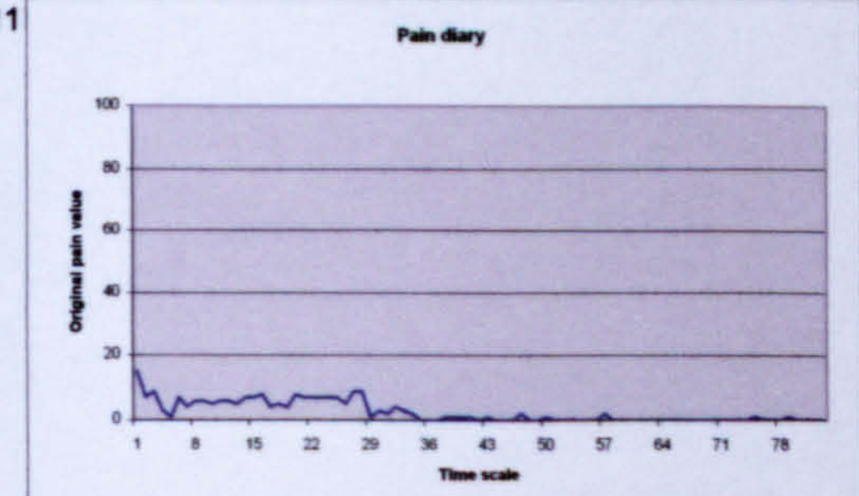
Pt No. 10



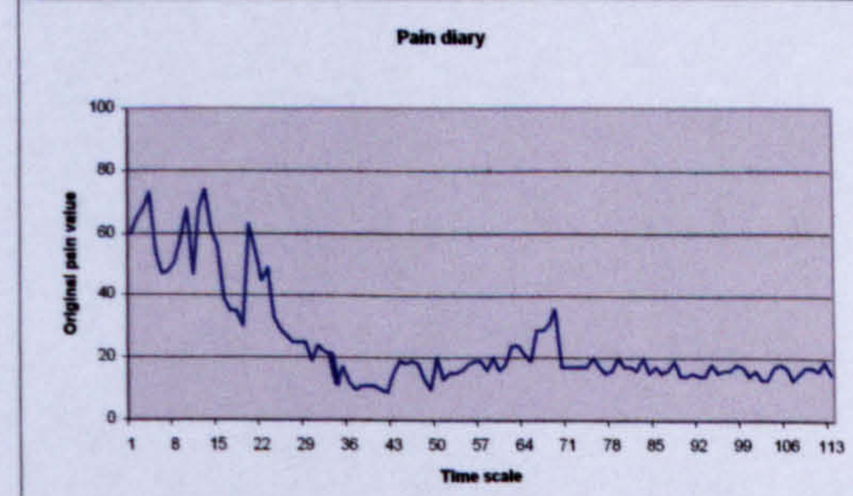
Pt No. 17



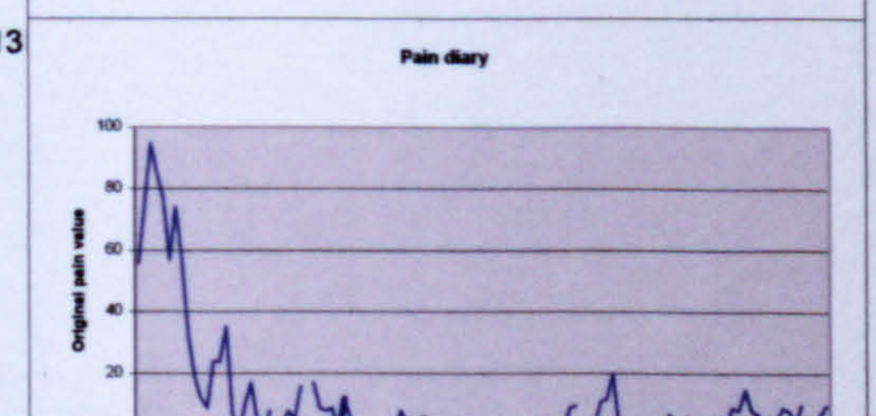
Pt No. 11



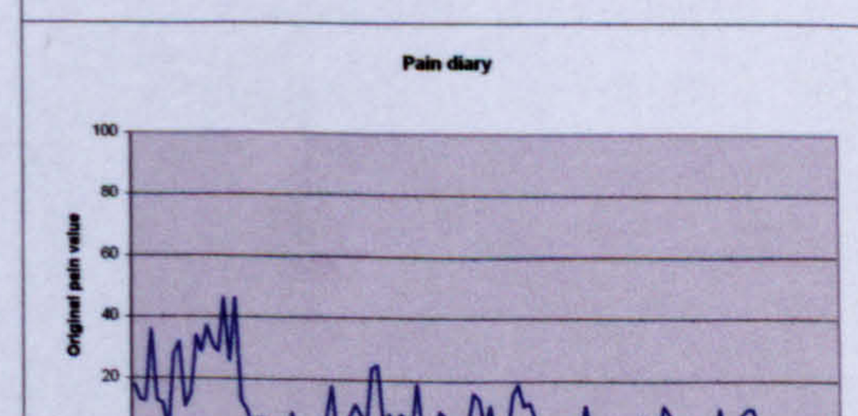
Pt No. 18



Pt No. 13



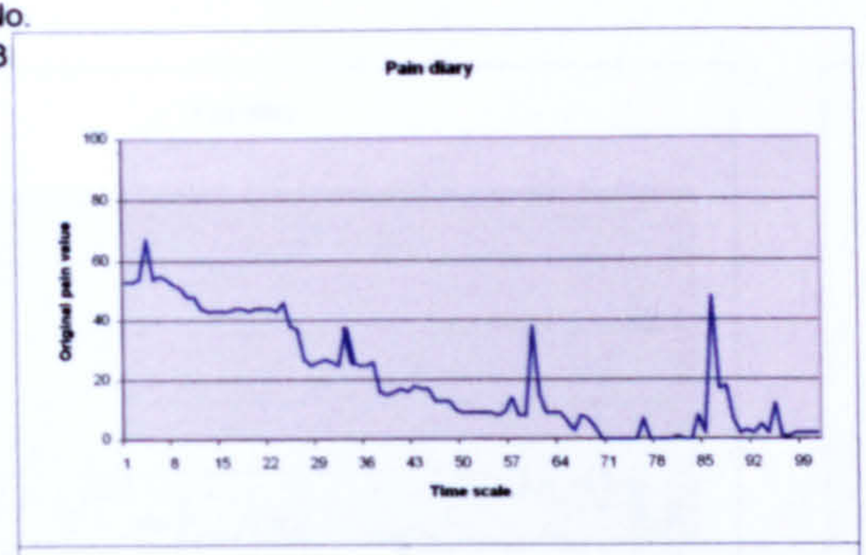
Pt No. 21



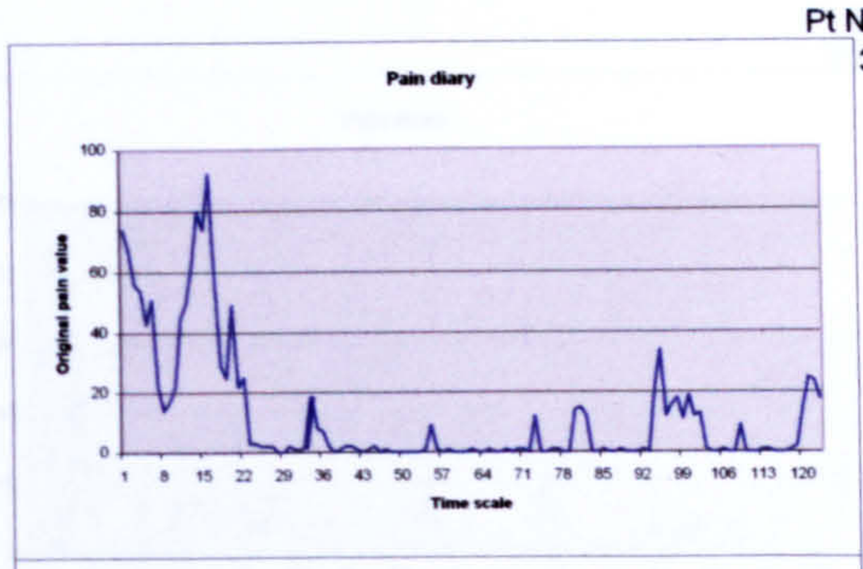
Improvers' charts

Page 2

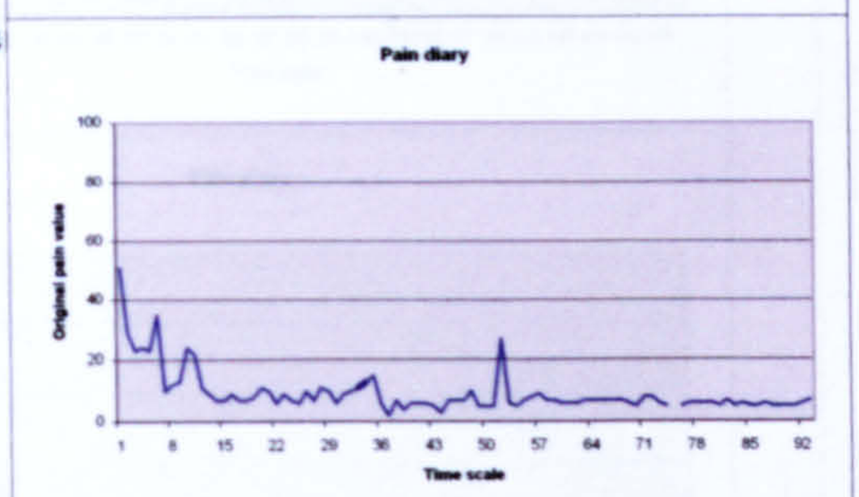
Pt No. 23



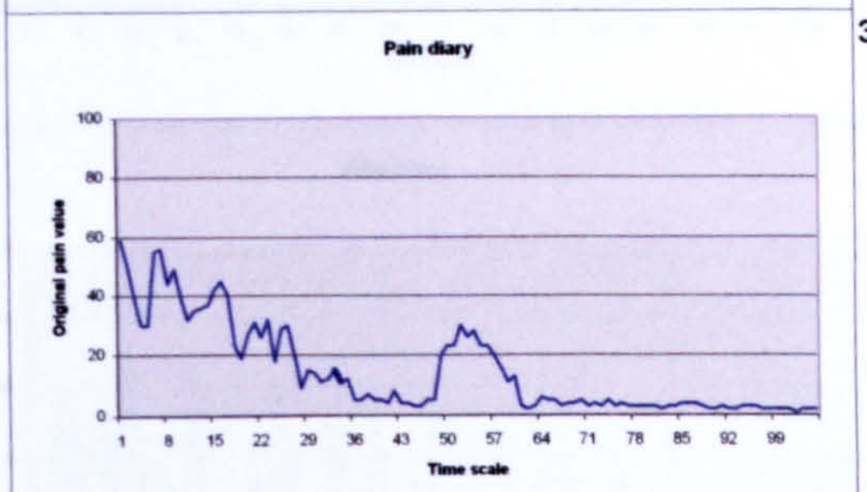
Pt No. 30



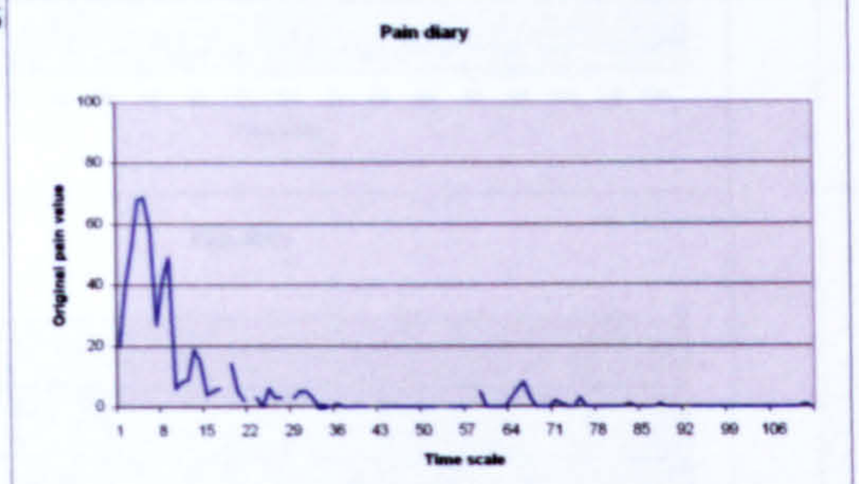
24



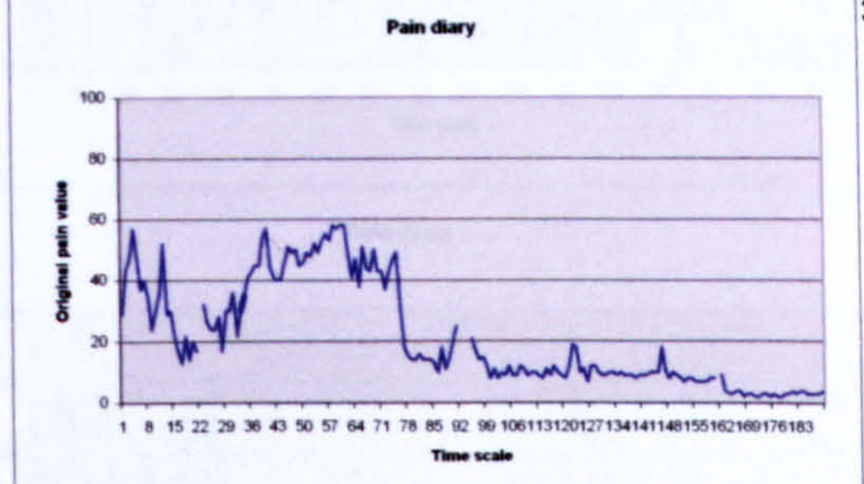
32



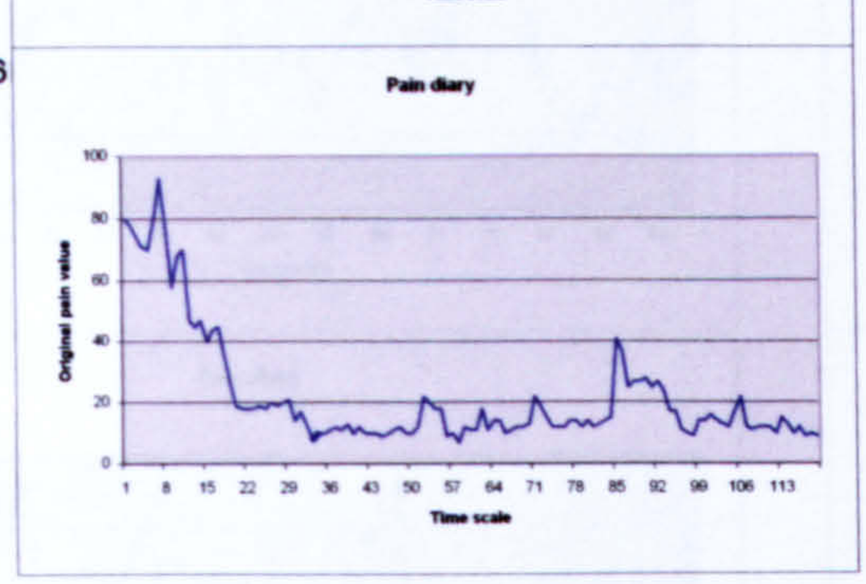
25



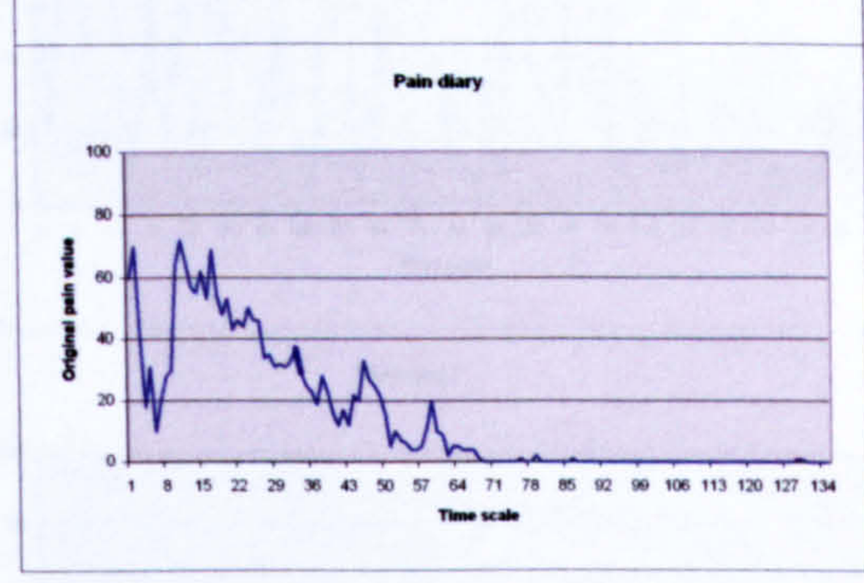
33



26

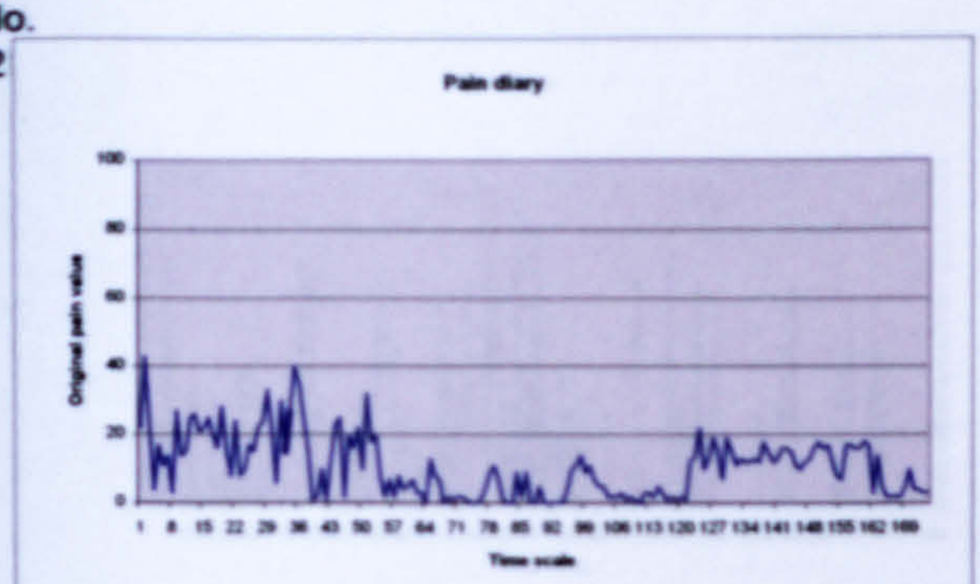


34

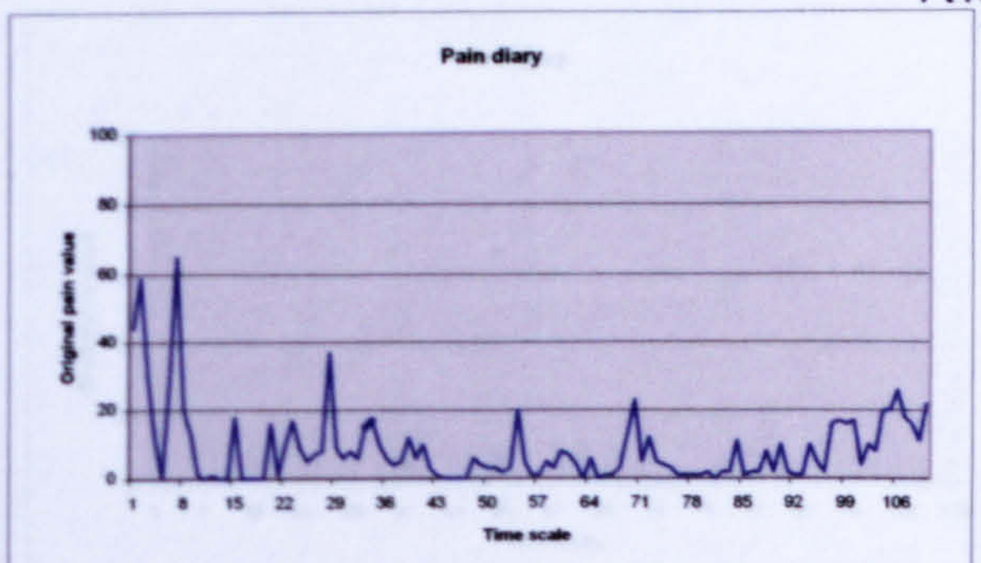


Borderline improvers' charts

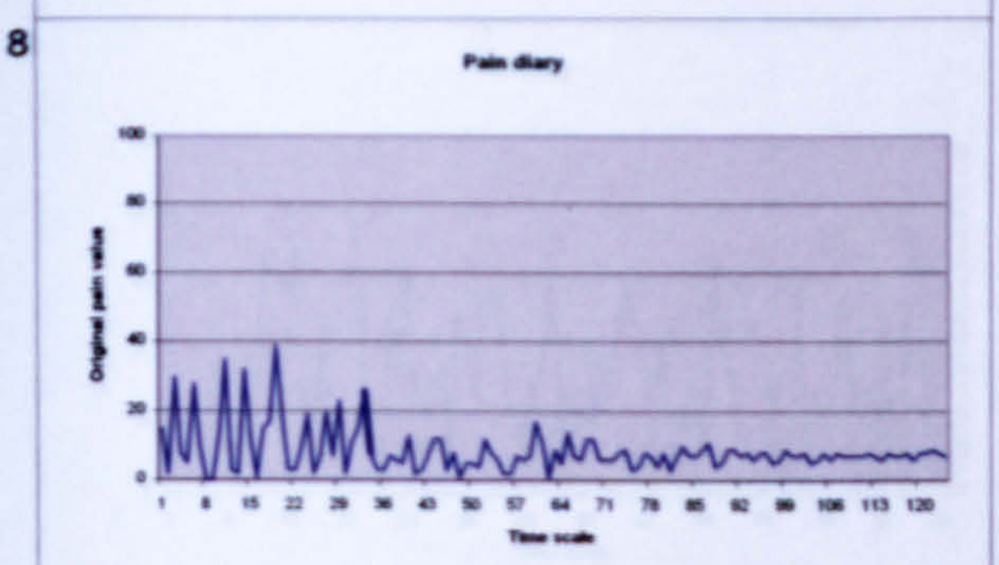
Pt No. 2



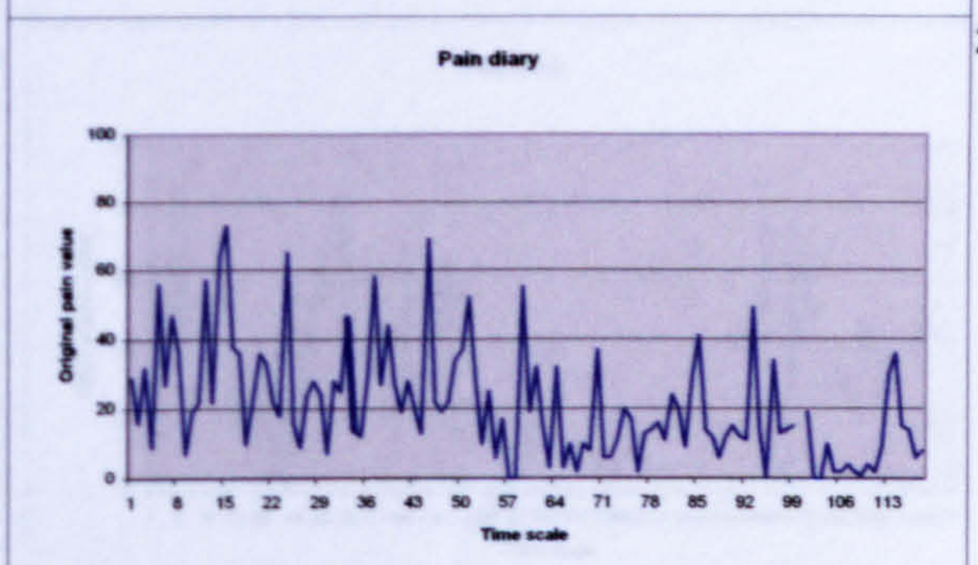
Pt No. 19



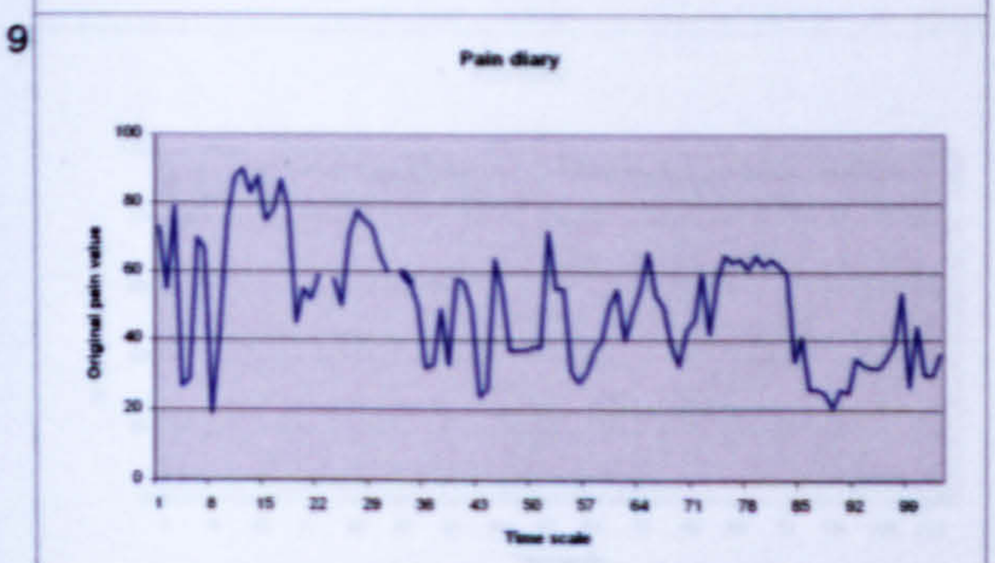
8



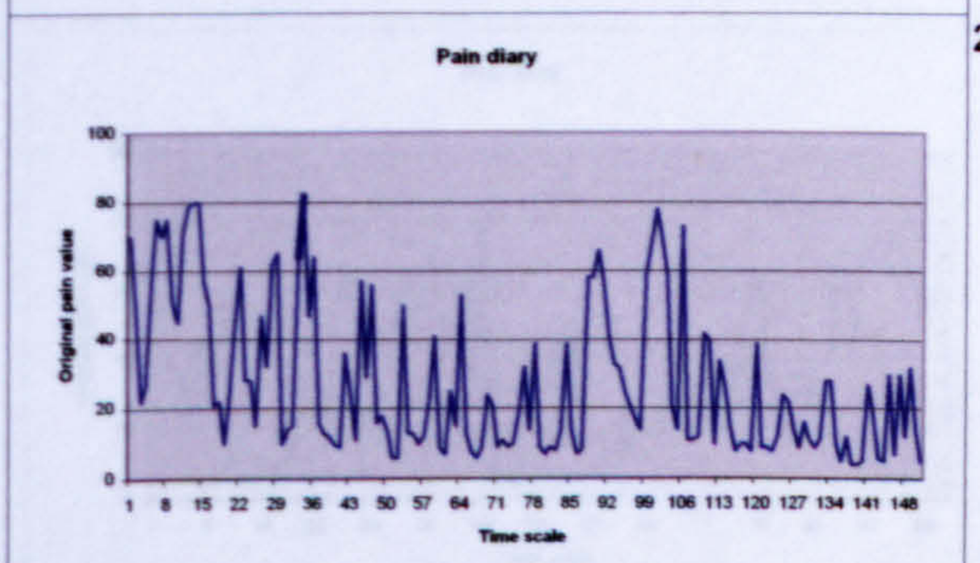
20



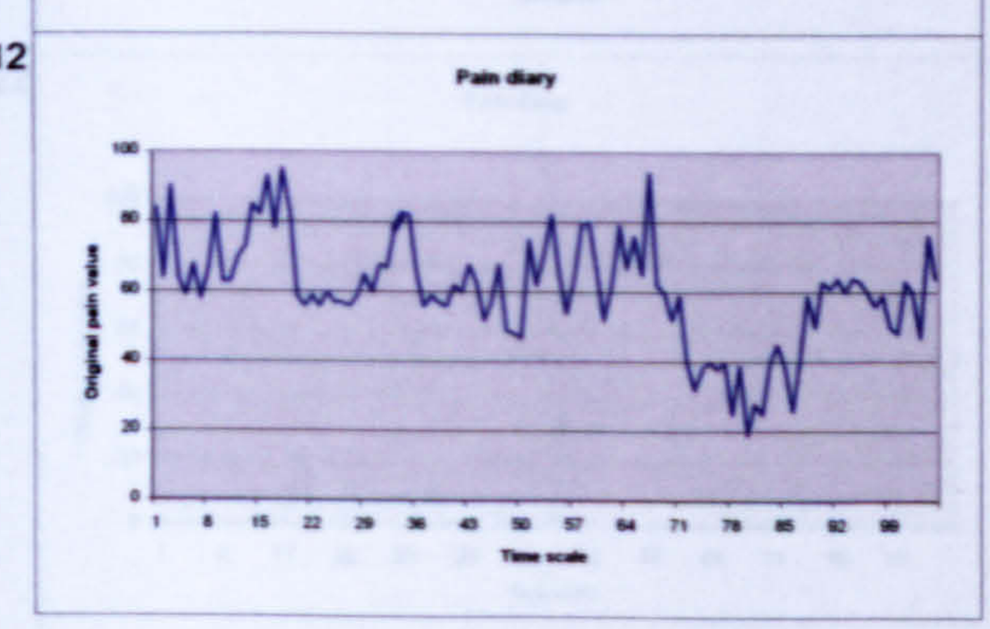
9



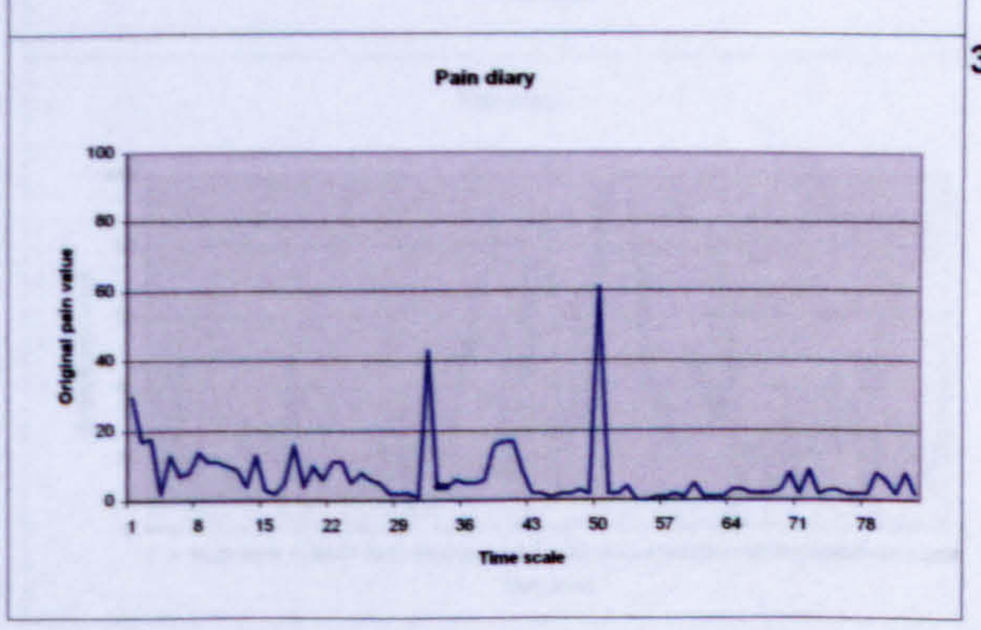
29



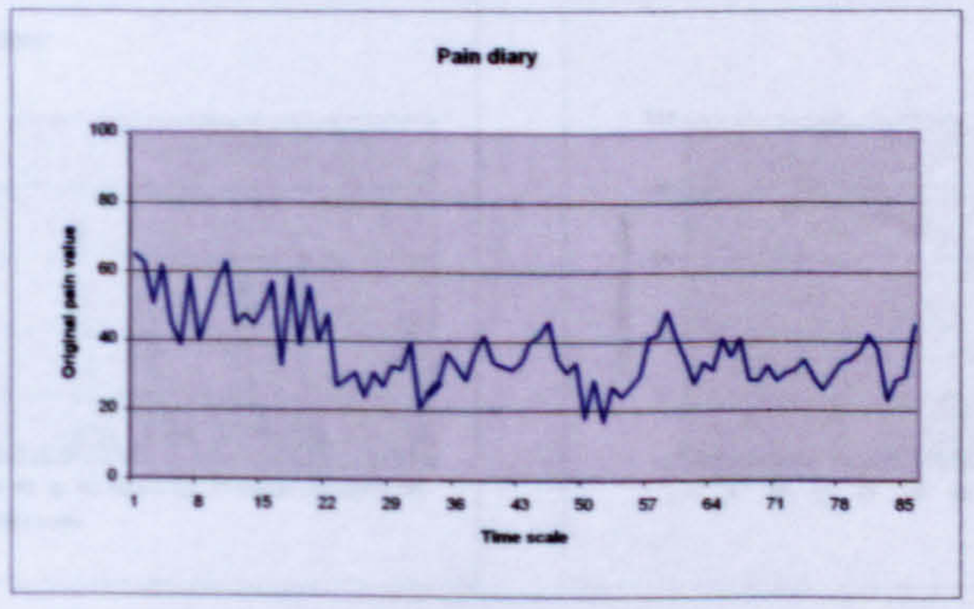
12



35

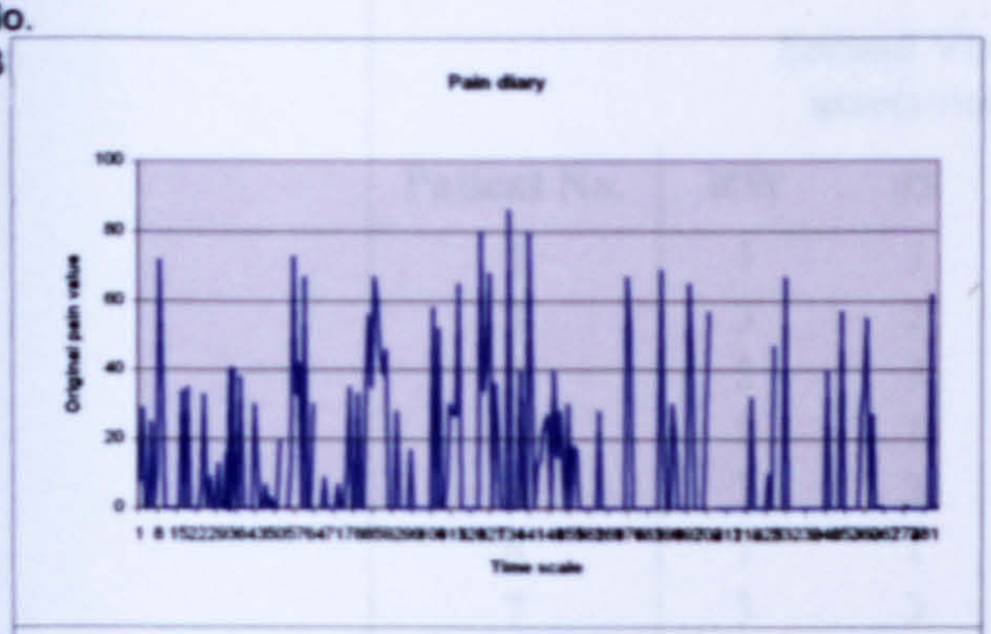


38

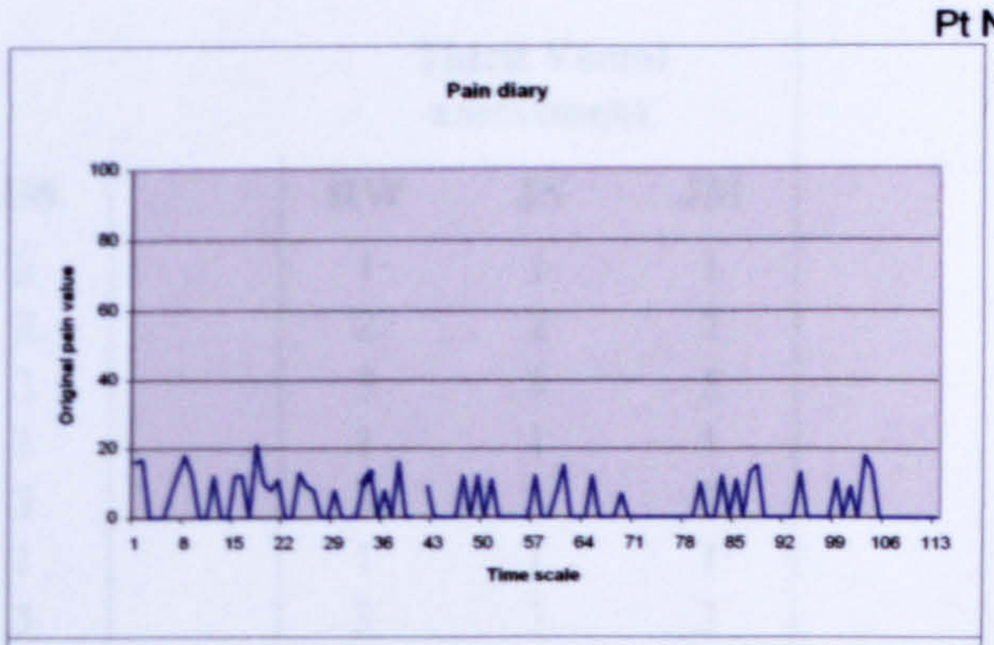


Non-improvers' charts

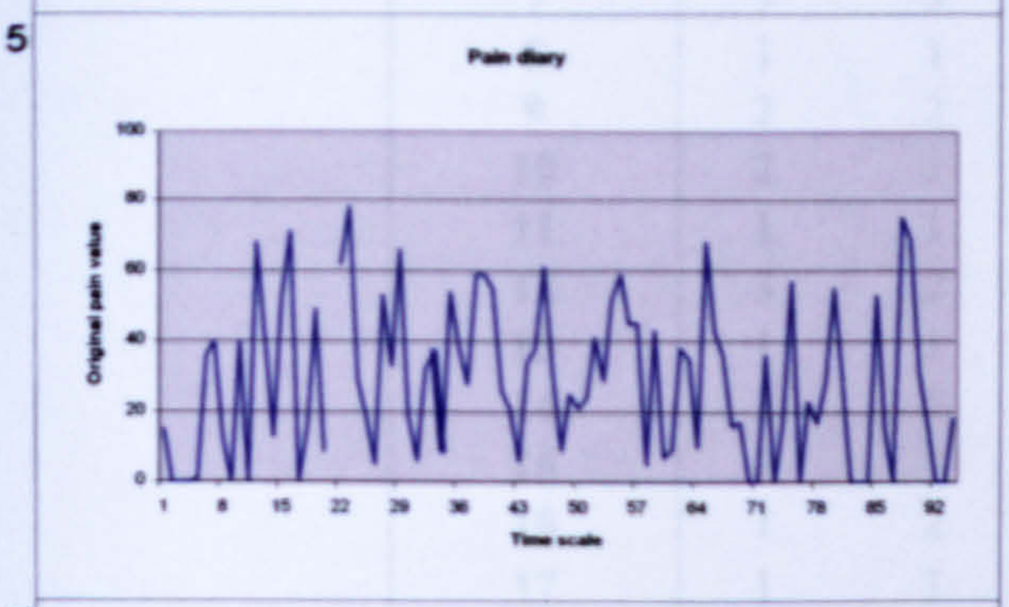
Pt No. 3



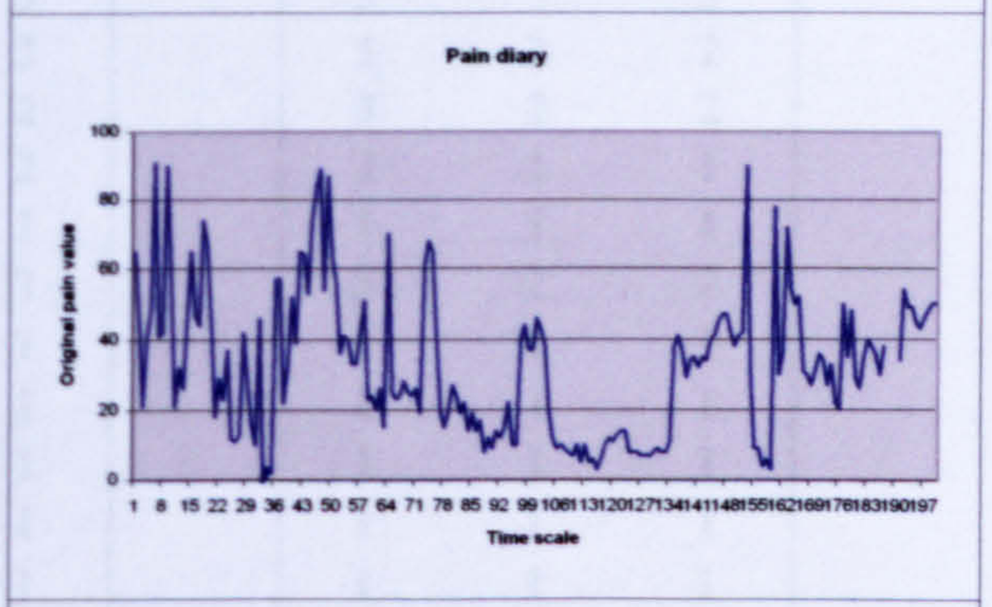
Pt No. 28



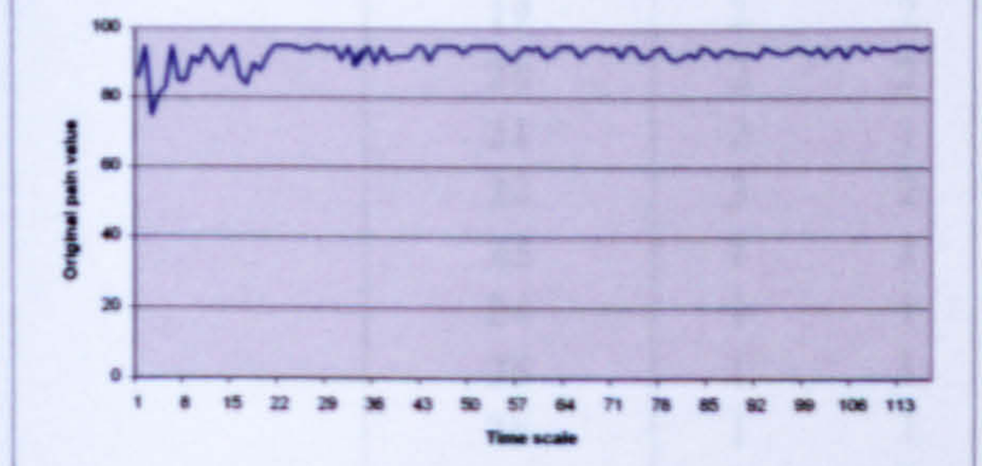
5



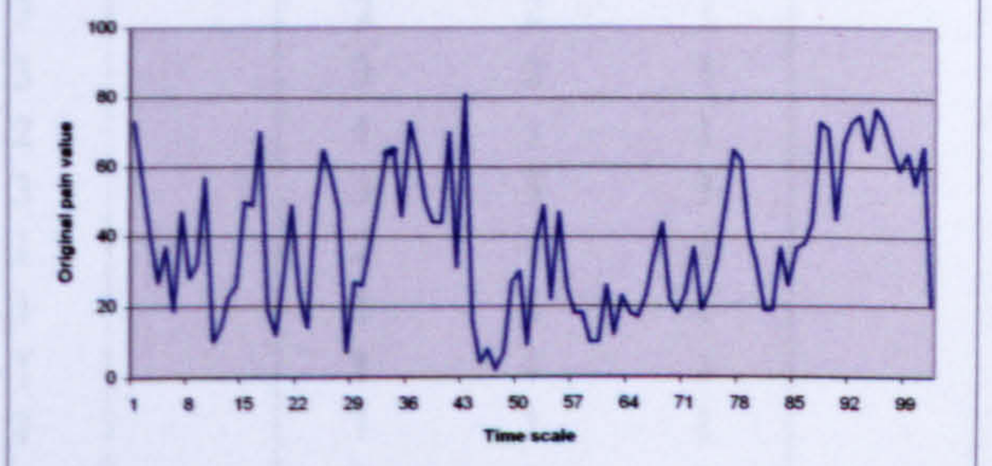
31



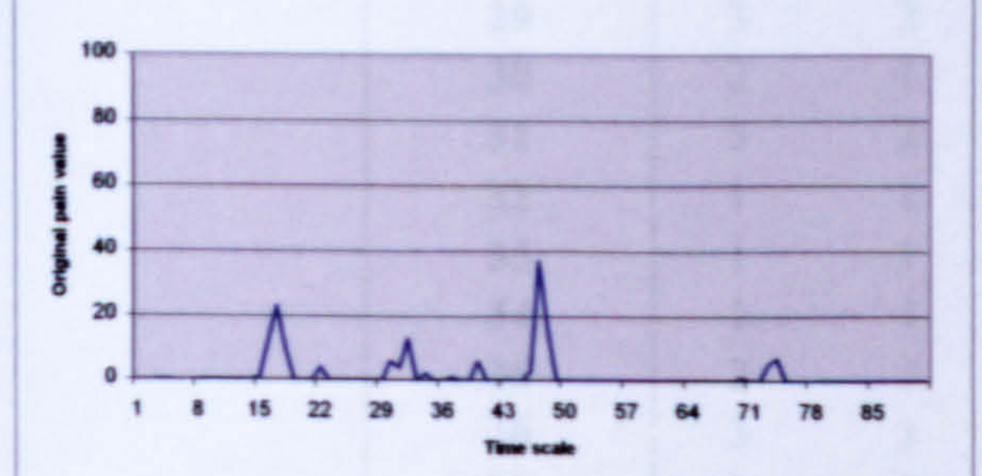
7



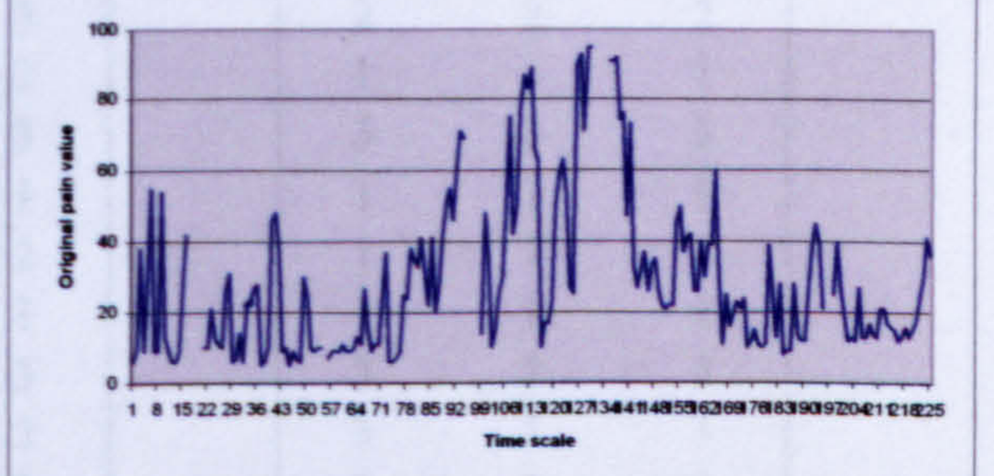
36



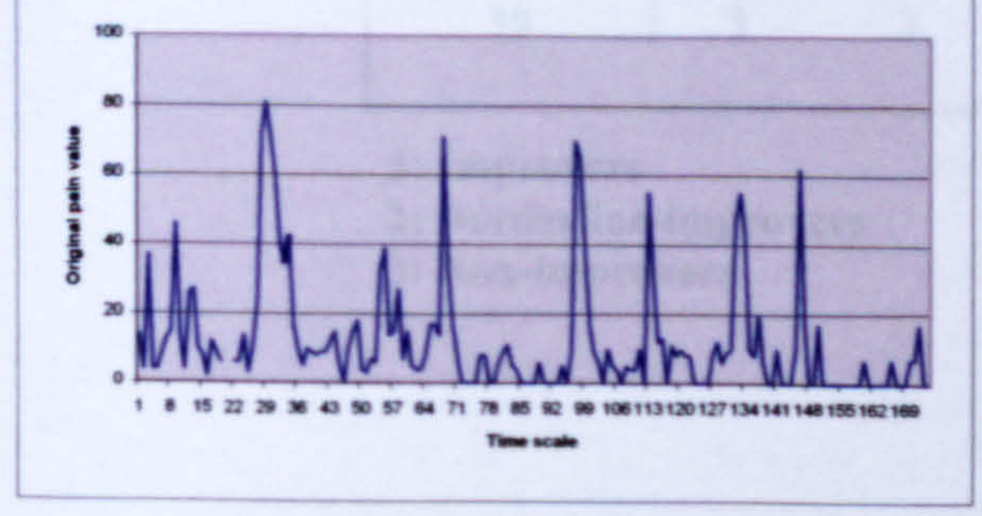
22



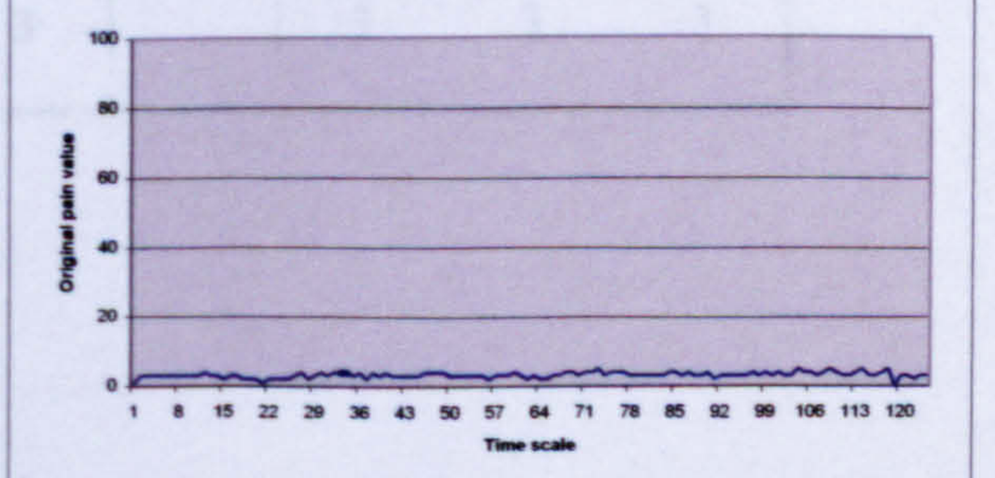
37



27



39



Professionals' visual assessment of individual pain/time plots

	Second Visual assessment			Third Visual assessment		
Patient No.	**RW**	**JS**	**JM**	**RW**	**JS**	**JM**
1	1	1	2	1	1	1
2	3	2	3	2	2	2
3	3	3	3	3	3	3
4	1	1	1	1	1	1
5	3	3	3	3	3	3
6	1	1	1	1	1	1
7	3	3	3	3	3	3
8	1	1	2	1	2	2
9	2	2	2	3	2	2
10	2	1	2	2	1	1
11	1	1	1	1	1	1
12	3	2	3	3	2	2
13	1	1	1	1	1	1
14	1	1	1	1	1	1
15	1	1	1	1	1	1
16	1	2	1	1	1	1
17	1	1	1	1	1	1
18	1	1	1	1	1	1
19	2	2	3	2	2	1
20	2	2	3	3	2	1
21	2	1	2	1	1	1
22	3	2	3	3	3	3
23	1	1	1	2	1	1
24	1	1	1	1	1	1
25	1	1	1	1	1	1
26	1	1	1	1	1	1
27	3	3	3	3	3	3
28	3	3	3	3	2	3
29	3	2	3	2	2	2
30	2	1	2	1	1	1
31	3	2	3	3	3	3
32	1	1	1	1	1	1
33	1	1	2	1	1	1
34	1	1	1	1	1	1
35	2	3	3	3	2	2
36	3	3	3	3	3	3
37	3	3	3	3	3	3
38	2	2	2	2	2	1
39	3	3	3	3	3	3

1: Improvers
2: Borderline-improvers
3: Non-improvers

Appendix 4
Participant information sheet and consent form used in *Study Two*

UNIVERSITY OF
NEWCASTLE UPON TYNE

Participant Information Sheet

DETERMINATION OF A SUITABLE REFERENCE PERIOD FOR ASSESSING TREATMENT OUTCOME OF TEMPOROMANDIBULAR DISORDERS (TMD)

"How your TMD symptoms changed over time and with treatment?"

You are being invited to take part in a research study, aimed to find out how your symptoms changed over time and with treatment. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What is the purpose of the study?
Temporomandibular disorders (TMD) are a set of uncomfortable and painful symptoms either in the jaw muscles, jaw joints or both. This study is designed to explore the changes in discomfort and pain levels in treated TMD patients.

What are the possible benefits of taking part?
It is hoped that the information that is gained from this study may give us a greater understanding of the issues of pain and discomfort and their measurement following TMD treatment. We are unable to pay a fee for your participation in this study as your attendance would be part of your routine treatment. We cannot guarantee that your participation in this study will be of direct benefit to you.

Why have I been chosen for the study?
Over the last few months, you have been treated in the Dental Hospital for the TMD problems you have been experiencing. You are invited now to take part in this study which aims to investigate your own experience of such symptoms and problems, and how they changed over time and with treatment. Altogether 20 patients will be invited to participate.

Do I have to take part?
It is up to you to decide whether or not to take part in this study. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive in any way.

What will happen to me if I take part and what do I have to do?
As a part of your routine review appointment, we would like to invite you, on this occasion, to an informal discussion with the researcher. The discussion will take no more than 30 minutes, and will take place in a quiet room next to the clinic. The researcher will tape record the discussion on a small portable tape machine. This is to make sure that we don't miss anything important.

Will my taking part in this study be kept confidential?
Your name will not be recorded on the tape, but we will make a record of your age and sex. Personal details and opinions you may express during the course of the discussion will be kept strictly confidential*. The tape recording will be kept in a secure manner in accordance with the provisions of the Data Protection Act (1998). The tape recordings will be stored in a locked cupboard with access only available to the researcher who has asked you to participate. The tape will be then transcribed and the recording will be destroyed within one year of the discussion. Afterwards, the transcriptions will be analysed, but will be coded so that we have no direct way of knowing which transcription belongs to which person.
The results of this study may be published in a scientific journal or presented at a research conference. In either case, your name will not be mentioned as part of the publication.

Who is organising and funding the research?
Who has reviewed the study?
This study has been organised by the School of Dental Sciences - Newcastle University, and is being carried out as part of a clinical PhD programme. It has been reviewed by Northumberland and Tyne & Wear Local Ethical Research Committee.

Contact for Further Information
If you have any questions please don't hesitate to ask either the interviewer or the consultant in charge of the clinic**. The clinic consultants will also deal with any clinical issues arising in the interview.

We would like finally, to thank you for the time you spent reading this sheet, and for participating in our study. It is much appreciated.

Version Number: 4
Date May 2005

---

* Confidentiality will be assured as with all research of this type, except in extreme circumstances where information is received relating to certain criminal activities, and all participants should be aware of this before they reveal possibly incriminating information.

** Drs RW Wassell and JG Steele, Senior Lecturers & Honorary Consultants. Department of Restorative Dentistry, The School of Dental Sciences, University of Newcastle upon Tyne, NE2 4BW
Tel 0191 222 7825 - Fax 0191 222 6137

Patient identification number for this study:.....
Ethics reference: 2003/38

CONSENT FORM

DETERMINATION OF A SUITABLE REFERENCE PERIOD

FOR ASSESSING TREATMENT OUTCOME OF TEMPOROMANDIBULAR DISORDERS (TMD)

Researchers: Mr. Adel Moufti, Mrs Jill Smith

Please initial box

1. I confirm that I have read and understand the information sheet Version 4, dated May 2005 for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that sections of any of my medical notes may be looked at by responsible individuals from the Dental School in Newcastle University or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records. ☐

4. I agree to take part in the above study ☐

Name of Patient	Date	Signature
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with hospital notes

Appendix 5
Topic guide used for the interviews in *Study Two*

Topic Guide for Qualitative Interviews

Introduction
Check the following
1. Adequate opportunity for participant to read information sheet
2. Consent form understood and signed
3. Consultant in charge is clearly established

"Thank you for agreeing to be interviewed, it really is very helpful. The purpose of the interview is to help us find out a little more about how the treatment for TMD works by asking people who have been through treatment, and listening to their answers. Because the information is often quite detailed and we need to pay very careful attention to what you say as you say it, for this reason we prefer to record the discussion so we can check on specific points when we need to. The tape will be destroyed when we have finished with it. We will not record your name or other personal details on the tape."

Overall history and change
Establish when symptoms started and when treatment was sought and started. Careful open questioning to gain an overall perspective on improvements or otherwise in condition?
Clarify timescale in as much detail as possible and probe further. Ensure the following are covered:
1. Why treatment has been sought (range of symptoms, effect on life?) Probe
2. Range of treatments provided and by whom, Probe and clarify as necessary
3. Which ones were considered to be effective? Why? (Briefly)
4. When each treatment started? Probe
5. When last treatment started?
6. What improvement occurred and when it was noted
   What would you consider improvement to be?
   How long it took to see any improvement – in reference to what time?
7. Was improvement gradual or sudden?
   Whether improvement is continuing (or not)
   When any improvement stopped
   Probe how improvement fluctuated
   Whether there were times when it got worse, when treatment went backwards

Quality of Life

1. Has improvement made any change to daily life?
2. Which problems motivated you to seek this treatment?
3. To what extent have your problems been solved?

Conclude

1. Would you like to add any thing?
2. Thank you very much indeed for your help

Appendix 6
Matrices used to map and summarise themes emerging from interviews in *Study Two*

First symptoms and main reasons for seeking treatment

	Summary of findings	Quotations
Participant 1	**Start:** 2 years ago - start was not clearly defined (a couple of years) **Reasons:** • Grinding and appearance awareness / confidence • Pain • Limited opening (at the dentist's)	*I would say a couple of years.* *I used to grind my teeth for a long time and I was conscious of doing it, and I had discomfort in my jaw.* *When I was having the* (dental) *treatment and I had to manoeuvre it* (the jaw) *around and when it went on and on that's when he* (the dentist) *decided that he could be doing any more and that's how he came to...*
Participant 2	**Started** 5 years ago, 1 month after a visit to a new dentist who made RCT which ended up with abscess and admission to hospital. **Reasons:** Pt linked it to an abscess	*It would have been probably after the abscess started on the first tooth she did which would been about a month, three weeks to a month.* *And me jaw locked on the one side, that's why I was admitted, taken into hospital. Very high temperature I had as well.*
Participant 3	**Start:** 1.5 year ago **Reasons:** Pt wasn't sure, related it to an RCT. Treatment sought when she started worrying about it and "felt sick of it". But she also had trauma at home at that time, and took anti-depressants	*I've had constant toothache - root canal work - tooth extracted* *It just sort of gradually got a bit better – then had started up again - it's been a lot better before I came anyway.* *Nothing you couldn't live with. But I was used to it* *I've never had any problem with my jaw ever* *It actually started when I got those anti-depressants it was just before not this Christmas past, the Christmas before that.*

Participant 4	Car crash 20 years ago, aggravated by oral surgery years later (7 years ago) She couldn't remember when she sought treatment, but this was about 3 years ago.	*It was quite a long time ago. I had a car crash when I was about 17 and I knew then that there was something that wasn't quite right. I knew I'd had some damage around there, but nothing was ever explained to me about that. But I think that it may have been aggravated further by treatment I had years later. So I actually had two wisdom teeth removed... They were deeply embedded in my lower jaw, since then I have had a lot more pain, a lot more discomfort, more headaches.*
Participant 5	**Start:** 6-7 years ago **Reasons:** Pain, then worries She had depression for so many years and is currently taking Amitriptyline.	*I can't really say when exactly, it's a few years back* *Clicks started 25 years ago, but actual problem started later* *Had post natal depression and father died of cancer / link to the symptoms wasn't clear* *I was always frightened this was going to ... and I think that's one of the reasons why I'm here... Well the pain plus the worry of wondering what was wrong with me. That really got me down.*
Participant 6	**Start:** 2 years ago **Reasons:** Pt couldn't link the start to any reason Was having the clicks before the ear infection but needed Tx only after the infection.	*I was having trouble with the ears, that was what prompted me to mention it to the Dentist* *I had been having the clicks for maybe about 18 months when I got the ear infection.*
Participant 7	**Start:** 1 + 3 years **Reasons:** clicking & locking jaw	*A few years ago* *One year suffer, then complained to dentist 3 years ago* *My jaw was clicking a lot, it felt like it was locking*

Participant 8	**Start:** 6 months after wisdom teeth extraction. Didn't inform her dentist straightaway because he said her teeth are alright. Informed her hygienist after 6 months. Received Tx from her, then was referred to NDH and started receiving Tx for 6 months now. Suffered for 2 years **Reasons:** Pain especially at bed time	*It was actually since I had my wisdom teeth taken out, that's when it really first started, I noticed it.* *And it was just a very slow decline really; deterioration in my mouth and the jawline really.*
Participant 9	**Start:** pain started after replacement of an old filling 1.5 y ago. Pt went back to dentist who denied responsibility. Pt received RCT at her private dentist's, then got the tooth removed. She visited a Max Fax unit with no useful treatment, then visited a pain clinic which couldn't help and referred her to NDH. **Reasons:** Pt blames the wrong Tx. Pain subsided after RCT but the bite changed. Occlusion adjusted but didn't help. Tooth extracted after 9m which eased the pain in the tooth area but the bite changed and jaw pain started.	*It started with dental treatment. I'd had a filling in one of my bottom teeth for quite a few years and I'd had the dentist since I was young and that filling had come out so I went to the dentist to basically to have it put back in and after that, after he refilled my tooth I started getting like pressure round the face and into my head... So I went back to my dentist and he said that he'd done all the treatment however it just seemed to me that it seemed so coincidental that it had come on since the treatment.* *he more or less told us that that wasn't' the case, that it wasn't anything to do with what he'd done, he checked me mouth and said go home, type thing, you must be like grinding your teeth or something.* *I had the tooth out in the November and, I don't know why, I don't know whether me jaw was damaged or what, but my bite never felt right. And from then on is when they started to take bits off certain teeth. And my dentist agreed with me that my bite wasn't right. But what he couldn't understand after that, and which amazes me, my bite changed to a point where it wasn't connecting in harmony with the rest of me mouth.* *My bite started to feel differently and in the January he* (the dentist) *did some work on my mouth where it corrected my bite and from then on the pain just knocked me off my feet. It went beyond bearability.*
Participant 10	Start: Pain started 2 years ago, and was treated at that time. Then occurred again 4-5w before coming to NDH. Pt was sent from his dentist to his GP who referred him 1st to a pain clinic, then to NDH. **Reasons:** Jaw pain and limited opening causing difficulty in receiving dental Tx	*I'd been to the dentist and they said they couldn't get in there to do anything.... I had like a sort of ruler in here to keep the jaw open so the doctor could get it* (the tooth) *out. Comical really. But it's definitely done the work there, you know.*

Range of symptoms

	Summary of findings	Quotations
Participant 1	• Grinding • Pain – Painful jaw (joint) – Was like an ache – Would take a while – Required painkillers (curative and preventive) – Worse in the morning • Clicking (tightness) (cracks) • Headaches – Dull headaches – Lasted for long time – Interfered with her daily life (but she was more concerned about her teeth appearance)	(The pain) *It would take a while. I used to take a lot of pain killers for my jaw and headache.* *In the morning when I used to wake up I felt tired and my mouth was tired.* (Clicking): *when I opened my mouth there was a tightness on the jaw and I used to feel that if I didn't manoeuvre it, it just cracked.* (Locking): *I used to have to wobble my jaw around before I could open it.* *I had headaches, never thinking the headache's from grinding your teeth at first.* (The headaches) *I would wake up with one in the morning and it would take quite a while to ease off and I used to also, if my teeth came in contact at any time during the day sometimes it used to trigger my headache because my teeth touched. ....And I've always been a good riser on the morning but I was very sluggish, you know, I would wake up with this headache...*
Participant 2	• Pain – Constant ache in the joint – Pain down RHS hand and body – Needed loads of painkillers • Headaches • Locking (is not well differentiated from DDWR)	(Pain): *I was in just so much pain; I didn't want to get out of bed.* *It's the worst ... when I spoke with the hospital it's the worst pain you can ever experience and I had two children and it's just the worst pain I could ever experience in me life, it was awful. Constant pain, 24 hours a day.* *Bad heads: Headaches, a lot of bad heads. Just constantly crying in pain all the time, very depressed* (Locking): *It did it quite a few times, kept locking on the one side, on the right hand side...* (and eventually) *it made its way back over the to same position.* *Oh I'd take 8 a day when it was very painful.*

Participant 3	• Intermittent headaches + painkillers • Pain in the joint & non-localized pain over the side of the face • Joint sounds (crepitus)	*I sometimes got headaches* *You can't localise pain again, it just was that side of my face.. I mean I knew if I touched here it was so sore* *if I yawned you could ... it was like grinding noise, it was always sort of like a jarring*
Participant 4	• Headaches and migraine – Migraine painkillers • Pain – Waking at night because of jaw pain – Discomfort, painful, sore, unpleasant ache – Aching jaw during eating – Periodical • Grinding • Clenching (Cramping / Clamping) at night and during driving her car • Locking last 4-5 months (later said one year) • Cracking once	*My GP referred me here and I went because I was getting headaches. I was waking up at night sometimes because me jaw felt so sore.* (Headaches): *I used to get a migraine. The headaches that I do get seem to be because I'm grinding my teeth and the jaw doesn't fit, I can feel that it's not right... I would certainly attribute the migraines to it as well where I would have to go somewhere dark and quiet.* (Pain): *I was waking up at night sometimes because me jaw felt so sore... There were times when it would be painful, but more often than not it would feel more like a really unpleasant ache, constantly.* (Grinding): *sometimes when I woke I would be aware that I had ... my jaw was clamping really tightly shut.* (during the day) *And I was aware that I was clamping me jaw tightly shut and stressing all that area around here* *The muscles would be tired – The muscles here always felt really tense. The muscles running down the front of me face always felt like they were strained* *Jaw felt so heavy and tired*

Participant 5	• Pain – constant (but was worse some days than the others) – pain even when talking or opening mouth – radiating to the neck • Clicking (was only "funny noise" 24 years ago), later became grinding • Headaches: permanent • Stiff jaw in the morning • "Seasonal affects" – Worse in summer – SAD in winter – Noticed the improvement almost since the clocks changed (in March)	*I couldn't even eat and I used to cry with the pain, it got that bad...it was awful.* (When there was no pain): *I just got on with it, I didn't really think nothing of it because I didn't have bad pain... You know, you don't think do you, if you're not in pain.* *I was in constant pain...and even talking, just opening my mouth* *In fact I had a permanent headache and my face was feeling as if it was going to split open with the pressure. Because it was the pressure on the eyes and it was really, really painful, that was what was the worst.* *I can remember some mornings I would get up and my jaw was so stiff it took me ages to get it working.* *It seemed to get more like grinding than clicking... but then I feel that it ... as if your bones are all being breaking.* (Seasonal affects): *Well I used to find that it was worse in summer time. But I also got allergic rhinitis that I got. So this could have been linked together you see. / I have Seasonal Affective Disorder. I sit in front of a lamp most of the winter.*
Participant 6	• Trouble with the ears (main complaint) • MFP • Clicking	*my cheeks were sore* *that side's not the same as that side. .just weaker or something* (The joint pain): *a shooting pain when it happens.. just like in me jaw like in my joints*
Participant 7	• Clicking • Periods of locking • Periods of constant pain (2-3 weeks) • Tight jaw • Day time, progress with the day	*I was clicking constantly.. just when I was talking or yawning* *It felt like it locked* *My jaw felt tight, it was like constant tooth ache in it, just a constant pain in my jaw.* *Only day time... get worse as the day went on* *it wasn't a constant problem that I had* *if my jaw locked I would maybe have say 2-3 weeks of like toothache and pain in my jaw*

Participant 8	• Pain – In the joint and muscles – Radiating toward the head, ears and eyes – Aggravated by talking and wide opening – The worst is at bed time – Strange description of pain!! (Hair, Strokes) • Swelling side of the face, unbearable to touch • Clicking in the "Jawline" on wide opening or yawning and closing • Difficulty on closing the mouth • Headaches (which she never had before) • Physical disability – Inability to brush her hair	(Pain) *It was really sore, and I couldn't stand anyone else to touch the side of my face.* *I felt as though it was at times, like the kind of nerves problem. You know, as if there was something attached to my face like a hair, you know, I was trying to take a hair off my face. It sounds really strange, but that's what it felt like.* *You can imagine when people are having Strokes their facial muscles all go, well that's what it felt like.* *I've never, ever had headaches in my life.* *My jawline was clicking* *I couldn't brush my hair without feeling very tender there.*
Participant 10	• Pain – Nagging pain – Terrible pain – Constant – Sore jaw – Periodic pain – Pain during the work can be ignored • Grinding	*Probably just discomfort because I find when I'm at work it's fine. Because I'm working, it's off your mind... probably because you're working it doesn't seem to bother you.* *Well it did affect my day to day life because, you know, I saw the doctor and the doctor said have time off work. But I'm not one for having time off work. To me, having time off work, is going to make it worse because it's on your mind 24/7. But I find, you know, if I'm actually at work, the pain's not there until I'm actually just trying to relax.*

Participant 9	• Pain – Pressure in/round the face – Pain into the head – Radiating (going down) across the (back of my) shoulders – Radiating down the neck, the arm, the leg – Pulling sensation – Extreme pain – Her face felt and looked paralysed because of the pain – Earache – Festering sensation – Strange description of pain!! (Festering, Stroke) – Not responding to any analgesics – Temporal characteristics: periodic, spiky, relapses • Non harmonious and unstable bite • Muscle spasm	*The only way I could describe it* (the pain) *was like somebody had tied something around your jaw and it was like a pulling sensation.* *Constant, agonising, burning, aching pain.* *It was just everywhere. It was just pain and the only way I can describe it is totally out of control.* *The pain was so debilitating - Totally, utterly unbearable.* *it's a pain that I never knew could exist. The type of pain even where there's times that I wished I was dead because the pain was just so unbearable.* *the pain was to a point where it started being all down one side of my face. To a point where the pain as that bad that my face was like paralysed... I felt I couldn't move it... I looked as though I'd had a stroke ... this was all because of the pain.* *The pain even got to the degree where it was like a festering sensation. I felt like I had earache. It festered right down the back of me ear, right down my neck, like a festering sensation in me neck* *The pain even got to that degree that it affected me right down into my arm... even into my fingers* *... right down into my left leg and all the way down so I was walking like I'd had a stroke.* *I mean they used to ask me to grade my pain between 0 and 10 and my pain was 10. It was agony.* *I had morphine patches and pethadine and ... which weren't touching the pain... I was on a lot of medication for my complaints, what I've been through, but nothing, nothing took the pain away. Nothing at all.* *The pain changes from time to time...it was having relapses where the pain would ease. It would ease down to maybe 4-5* (on the VAS) *and then within a couple of days it would shoot back up to about 10.* *It* (the pain) *seems to go through days where I might have a good couple of weeks or a good odd day. And then all of a sudden whoosh. The pain will ... I'll have like, what I call, a relapse.* *The pain can vary daily or it can actually vary from getting up in the morning to sort of mid-morning to mid-afternoon. The pain can vary throughout the day.* *My bite changed to a point where it wasn't connecting in harmony with the rest of me mouth... My bite never felt as though it had ...there was no stability within my mouth.* *my face spasmed. I had a constant face spasm because the muscles were tort and tight and my jaw... was twisted and distorted and tight. I could actually feel my muscle spasms in my face.*

Changes in oral function

	Summary of findings	Quotations
Participant 1	• Difficulty chewing meat and food needing long time of chewing but this didn't affect her. • Yawning (crack) • Problems with dental treatment (opening for long time)	*I had difficulty chewing meat. Anything that I have to chew for any length ...* *I used to have to just pick, you know, pick certain foods that I didn't need to chew because my jaw used to click quite a bit...* (But) *it just wasn't worth the hassle, you know, it was an easy option, not to eat meat.* *When I yawned or opened my mouth quickly it was a crack.* *When I used to have* (dental) *treatment I had difficulty sometimes to keep my mouth open for any length of time.*
Partic. 2	• Couldn't eat anything but soup • Even not able to open mouth • She lost weight	*I couldn't eat nothing. I had to eat soup and drink pop all day ... I couldn't hardly open me mouth.*
Partic. 3	• Avoid chewing on the painful side	*I'd been chewing on that side, which I never used to do, and I still don't do very often, but I had forgotten about it.* *If I yawned you could ... it was like grinding, that's what it felt like, a grinding noise*
Partic. 4	• Aching jaw during eating • Affected eating to some degree	*Only when it was really bad and you couldn't eat... There were some occasions where I was, you know, you could just have soup, soft stuff. When it was really, really bad but I can only remember that ever happening twice, actually, to that extent, where you were unable to eat anything.*
Participant 5	• Affected her eating • Precautious when yawning • Avoid wide opening	*I couldn't even eat...* *I mean I never could eat hard things, chew things. I used to have my meat really cooked softly. Hard sweets, things like that I never used to eat.* *If I yawn I tend to try to keep my mouth as normal as possible* *I can't open my mouth wide* *I was in constant pain. and even talking, just opening my mouth was really ...*

Partic. 6	• No impact – patient is just aware of her eating	*Well I'm aware of what I eat.. Opening my jaw for an apple or something is a bit, you know, If I want to get a bigger bite.*
Participant 7	• Avoided hard food like apples • Was told to stop chewing gum	*I've stopped eating hard food like apples and that because it was just aggravating it more* *Didn't affect Chewing meat* *They advised us to stop eating chewing gum. I found that a big help actually* *When I was yawning I was clicking*
Partic. 8	• Pain and clicking on yawning and opening wide and closing • Talking was painful • Difficulty on closing the mouth	*I found that if I'd been talking a long time, speaking a long time on the telephone it was really painful... I was I was really upset about it, frightened more than anything because I felt as though I wasn't able to speak properly.* *I couldn't close my mouth properly at all and I was quite frightened.*
Partic. 9	• Pain affected talking and walking	*Altered the way I talked, altered the way I walk, everything* *Just like it* (the pain) *has paralysed you in a way you just move your arm without thinking, I was more having to use my right arm to lift my left arm to move around.*
Partic. 10	• Limited opening causing difficulty in receiving dental Tx but he was not AWARE of his problem with mouth opening	*And then I found when I had to go to the dentist ... she couldn't get in to do any treatment. She couldn't get in me jaw.*

Impact on quality of life

	Summary of findings	Quotations
Participant 1	She was too concerned about the effect of the grinding. Headaches interfered with her daily life	*My teeth were changing and it was changing my whole face really... and my appearance.* *I always did take care of my teeth. I used to have nice teeth and the deterioration from that was drastic.* *When I looked in the mirror it sometimes doesn't reflect me. It changed my appearance so much.* *My confidence because I had what ... well I couldn't smile with grinding my teeth the smile is only a part of it, you know, and when I look at photographs of when I was younger, there's no comparison.* *(The headaches) They had interfered with it (daily life).* *I've always been a good riser on the morning but I was very sluggish, you know, I would wake up with this headache...*
Participant 2	• Impact on mood – depression – needed anti-depressants • affected her daily life • stopped her sleeping • linked pain to confidence • considered this pain worse than labour	*Headaches, a lot of bad heads. Just constantly crying in pain all the time, very depressed ...Just really down, I felt down all the time... Oh very, very depressed. They put us, when I went to the hospital, it was Diazepam or something which was sort of a painkiller but also anti-depressant as well she told us.* *I just wanted to lie in bed all the time I was in just so much pain, I didn't want to get out of bed... I just crying all the time.* *Well you feel a lot more confident in yourself when you're not in pain and ...It really gets you down when you're in pain like that, it's excruciating pain, just like constant toothache, 24 hours a day all the time.* *just to check to talk to somebody who ... because I mean you talk to people and they don't understand what you've got. It is a very painful thing; it's not a very nice thing to go through and, I mean, they under the case when I come.*

Participant 3	• No impact on daily life • No impact on sleeping quality but physical impact (couldn't sleep on her side, but wasn't worry about it) • Worried about her conditions • Took anti-depressants (but it was not clear whether this was for her TMD or trauma at home)	*Well it didn't stop me doing anything.. I still was working* *something you learn to live with* *It did affect you sleeping - , I think more difficult really - It wasn't all the time- but it wasn't anything I could say I'd be worrying about.* *you had things to occupy you all the time (day time), you didn't have time really to think* *I was really depressed; I can tell you about it. I was really sick of it. I felt right down with it. I think that's, that was the reason, that was the finish I thought.* *I was starting to worry about it, I was imaging everything. Because that's the type of person I am.* *When I got those anti-depressants it was just before ...*
Participant 4	• Distress • Feels low • Affected her daily life • Affected her sleep for some time • Made her feel unhappy • She is a social worker and she has a "quite stressful job" • Improvement was noticed in terms of what she feels not what she does • With Tx she felt more confident	*It was really distressing* *It did make you feel quite low because it would happen in periods. So it was certainly affecting me in that respect* (everyday life) *and yeah making me feel really just generally unhappy.* *How I feel. Not so much what I do, but how I feel.* *Because my colleagues give over laughing at me.* *A bit more confident that I'm not going to just feel like the headache's coming on at any minute* *It might happen for two, three, four weeks where it would be particularly noticeable for me. I'd be waking up, I'd have headaches, I'm locked* *My GP referred me here and I went because I was getting headaches. I was waking up at night sometimes because me jaw felt so sore.*
Participant 6	• No impact on QOL • No impact on sleeping but painful jaws on waking up	*No it wasn't affecting my daily life.* *When I was waking up in the morning my teeth bones were sore, you know.*

Participant 5	• Cry with pain • Was worried about her condition – this got her down. • Wasn't sure if it affected her daily life • Affected her sleep • Thought her well being is linked to the improvement.	*I used to cry with the pain, it got that bad.* *Well the pain plus the worry of wondering what was wrong with me. That really got me down...I suffered with depression with it.* *It improved with my well being as well. I'd get a sort of glow... some days to cope with are better than others.* *I used to wake up with neck pain* *If I'm feeling good, on a good day everything doesn't seem as bad.* *I can have a day where I feel that I want to sit and cry all day. And then I can have a day feeling as if I'm on top of the world sort of thing. I'm very unpredictable.*
Participant 7	• Nothing mentioned about everyday life • Mood altered with pain • No impact on sleeping	*I would say it would alter my mood when the pain was there.* *It made me quite down* *I would have been snappy*
Participant 8	• Affected her mood • Distressed • Upset • Frightened • Scared to speak. • Worried from headaches which she never had before • Silent suffer: Didn't tell her husband , nor the dentist • Worries of cancer • Affect on her work, and asked to be moved from her job. • Believed some of her feelings were of psychological reasons	*Yes it did, certainly did distress me* *I was really upset about it, frightened more than anything because I felt as though I wasn't able to speak properly.* *I couldn't close my mouth properly at all and I was quite frightened. I didn't say this to my husband.* *I was a bit scared to speak.* *I've never, ever had headaches in my life... That's why I was worried.* *I had x-rays and so forth and he* (the consultant at NDH) *said what the problem ... there was obviously no real sort of like cancer or anything like that which I was thinking... I was relieved that it wasn't tumour growing...when you feel a lump in the side..* *I worked as a receptionist and I found that if I'd been talking a long time, speaking a long time on the telephone it was really painful by the end of the evening, it was really sore...and I actually asked if I could move off the switchboard. I was that concerned about having to speak for a long period of time.* *I think a lot of them might have been psychological as well*

Participant 9	• Had to leave work for a year • Severe impact on everyday life: everyday activities, sleep (see function), concentration • Had to move to her parents' and to rely on them for everyday activities and to look after her child. • Was confined in her world • Sever impact of pain • Despair • Affected her mood (get her down) • Impact on sleep	*It affected my daily life. It made me feel just poorly, unwell.* *It stopped me doing everything. I've been off work for over a year.* *I was totally reliant on other people for my everyday things. Getting dressed, getting bathed, the pain was that bad. My mum was having to put me in the bath and wash me and shower me.* *I was having to live with me mum and dad because the pain was that bad I couldn't look after my little boy. I couldn't look after myself.* *I couldn't even do shopping. I was confined to the house really... you know, you're just sort of in your own little world with the pain because it's just so bad.* *If I had to concentrate on anything my concentration span ... like I was sat down with my little boy on the floor, I couldn't concentrate on doing like putting, helping him do a jigsaw, anything like that was just ... too much for me, yeah.* *The type of pain even where there's times that I wished I was dead because the pain was just so unbearable.* *I mean they used to ask me to grade my pain between 0 and 10 and my pain was 10. It was agony.* *The pain was so debilitating, it totally controlled my whole life.* *The only thing I can say to you is I never thought the day would come... where I don't appear to have no pain..* *Despite having fantastic little boy I just wished I was dead because I can't cope with this intensity of pain any more.* *Sleep was impossible. The pain was so intense. I think the only time I went to sleep was when I think my body was just so run down and knackered with the pain I could sort of cat nap.*
Participant 10	• Contradiction between reporting no impact on daily life, and an impact on his work • Difficulty in sleeping	*Do you feel it's had (the pain) any effect on your daily life, the things you do, how you do them? -- No, not really, no.* *Right, okay. So it's never been as bad so it would stop you doing something? -- No. No way.* *I think I had the problem at work because me jaw was sore, me teeth were going absolutely ...* *I can't really sleep on this side because it's just an ache. So I ... I tend to sleep on that side anyway.*

	Summary of findings	Quotations
Participant 1	• Very much improvement • Definition of being better • Improvement as reduction in medication • Importance of adhering to treatment • Hard Splint – was desperate to do the treatment. Was positive. It made her look better!!! and feel better!!!	*Very much improvement... I don't have the pain at all and my jaw doesn't click, you know, I can open my mouth freely, you know, and I don't live on painkillers.* *I used to take them every 4 hours. Even if I weren't particularly in pain I was frightened that it would come back. It was preventative.* *The more you kept it in it shortened the time ... the longer you kept it in it shortened your time. So the treatment started to getting better, you know. And somebody that maybe kept it in a couple of hours a day ... but I really was desperate to start the treatment* *I used to wear my splint all day at all times. Even that made me feel better even with having the splint in...It made us look better.* (The splint) *It stopped everything really.. It did because it, you know, my teeth didn't touch. It's amazing how it's made, you know, it was just ... I could open my mouth freely with it.*

Participant 2	• Feeling better in herself • Reduction (but not disappearance) of pain • Felt more confident • Improvement as reduction in medication • Improvement started after physiotherapy straightaway but was gradual over 8-12 w. • Improvement didn't stop or reversed • The locking stopped "very fast" while the headaches and the pain reduced gradually • Improvement stopped now and can't be better • Coping • The use of treatment • Was advised to avoid chewing gum and hard food but this didn't bother her. • Sequence of Tx and starting time was difficult to remember	*Well I feel a lot more better in me self and it's not as much pain either as what I was experiencing... definitely not as much as when I used to get it.* *Well you feel a lot more confident in yourself when you're not in pain.* *I feel very confident now since I've been here, a lot more confident now.* (Painkillers): *Oh I'd take 8 a day when it was very painful..(but now) Not half as much, not hardly any actually.* *I used to go* (to the physiotherapist) *once a week so I would say about 8-12.* *It's a lot better. I think it's the best that it's going to get now I don't ... it's alright, I mean I feel that it's okay now, I do, yeah... It's not going to get any better.* *Now I can cope with it a lot better.*
Participant 3	• Pain, headaches disappeared • Crepitus is better but couldn't remember when it disappeared • Improvement to QOL • She didn't use the splints straight away, but when she did she noticed next day improvement (pain) • Took a month to notice any improvement • Took a month to notice reduction of headaches • Improvement was constant and gradual – no relapse	(About change to your daily life): *Oh well certainly* *When I have had any ache in at all I've used the splint. And it has been, you know, much better the next day.* (It took) *a month probably* (to notice any improvement) (Crepitus) *I thought of this last week, I thought just for some reason that doesn't sound as bad, because I'd been chewing on that side.. It's because I'm just ...* *It's gradually improving, it just seems to be going on one level now apart from the odd thing...the treatment never got worse*

Participant 4	• Reduction of headaches • Improvement as reduction in medication • stopping her grinding & clamping • Jaw locking stopped • Improvements impacted on how she feels but not what she does • Feeling better, and more confident • Initially couldn't remember how long it took to notice improvement, later said it is a fairly quick improvement.. an immediate relieve • Some symptoms relieved quicker than the others: the grinding (and the locking) immediately, the headaches a couple of weeks later (noticed by her partner before herself) – later said these were gradually improved. (It's just a day to day thing) • Improvement stopped and reversed. New symptoms developed.	*It just became noticeable that I wasn't waking up with like really nasty pains or headaches* *To be honest* it (the improvement) *was fairly quickly because it was stopping me from that hard impact grinding me teeth and clamping through the day.* *I think some of the things I was ... that were happening to me were relieved immediately.* (Headaches): *In fact my partner began to notice I think before me. He mentioned something just after a couple of weeks or so.* *I found that it ... either my jaw had changed shape or the clamping and grinding had altered the shape of the splint. It wasn't fitting properly it was ... and I was again getting headaches more frequently and then the strange clamping thing started. You see I had not came back. I'm positive I was sent an appointment out after all this and I didn't come back for that...Because I was a bit worried about anything else being done after all those years.* *I did notice that certainly through the day the clamping was ... because it was giving it was quite a thick routine really, it was a spongy, I wasn't getting that, this cramping feeling.* (Effectiveness of the splint): *When I take it out I do notice that me teeth sit differently in that they don't fit. And I think ... I don't know if that's because I'm more consciously closing me jaw straight instead of moving it.*

Participant 5	• The splint reduced the joint stiffness and improved her mood. • The splint didn't help later and there was a relapse. • The Amitriptyline helped her muscle to relax. • Improvement was gradual • Clicking became grinding and didn't disappear. • No improvement to function • Improvement is not continuing / achieved only 50% improvement • Noticed improvement 6 weeks ago, didn't know why. Later linked it to the clock change (seasons affect) • Coping ability	*I was waking up feeling better. You know, when you wake up each morning to it, some days you don't want to get out of bed.* *In the morning when I woke up I didn't feel as if my jaw was as stiff.* *The splint helped and in the morning I used to feel better, and it just made me feel better in my moods.* (The Amitriptyline): *it's since I've been on the Amitriptyline I've felt as if I've got more movement in my face...I feel better...My face feels more relaxed. There's been an improvement across my face. I would say top jaw has an improvement but I still have stiffness and pain with the bottom.* *I think it improved with my well being as well. I'd get a sort of glow... some days to cope with are better than others.* *The splint took 6 months to notice an improvement, just when it was adjusted.* *Well I still don't eat solid like hard chewy or toffee or boiled.* *I wouldn't say I'm feeling better every day, but I'm feeling that what I've got I can cope with better...I wouldn't say my problems have been completely solved; I live with them better some days than I do others.* *Well I feel 50% better than what I did feel.* *What I've got I can cope with better*
Participant 6	• Needed 10 days to notice changes • Information about improvement were excluded as the patient didn't use the prescribed treatment, and although improved spontaneously she was in a relapse period when was interviewed.	

Participant 7	• The clicks have gone gradually • Relief of pain within 3-4 days with the treatment (2-3 weeks without it) • Locking eased • it took 2-3 weeks "to clear up" / gradually • Tightness eased • Perceived improvement: – less time is needed to obtain the relief – Recurrence of the problem is less frequent • Still avoid opening wide • Relapse happened after about 3 Y • Coping ability • The problems weren't cured • Relief from the worries	*Before I came here if my jaw locked I would maybe have say 2-3 weeks of like toothache and pain in my jaw. But once they explained about the exercise I would say maybe about 3-4 days down the line I would start to feel ...* (Before Tx): (The problem) *would go away and it would be okay for a month. But once they showed me the* (exercises)*... I would maybe go maybe a couple of months again where I was fine.* *It took about 2-3 weeks to actually clear up... to get rid of the ache completely.* *I would say maybe about 3-4 days down the line I would start to feel ...once it cleared I would maybe go maybe a couple of months again where I was fine. But if I started feeling any pain again I just got straight on with the exercises.* (Function): *I Still don't eat apples that's something ... I'm frightened ... if I do eat apples I'll cut them, I won't bite into them.* *If I started feeling any pain again I just got straight on with the exercises... until it was clear.* *I would say* (my problems) *solved completely. Not actually cured because it still can come back, but I know how to deal with it now.* *It's the relief ... because you get it in your mind that it's something...*
Participant 8	• Immediate relief • Disappearance of pain and reduction of headaches • Improvement as reduction in medication • Relaxation of the jaw and disappearance of the swelling • Ability to talk easily • At work: the handset doesn't hurt any more • Reduction but not disappearance of the clicking • Improvement to the feelings and mood, both at work and home.	*I was relieved that it wasn't tumour growing* *And I feel as though I can talk quite easily, you know, quite freely because I was a bit scared to speak.* *Happy at work because I don't have the pain and just generally happier altogether.*

Participant 9	• Complete disappearance of the facial pain, and localization of pain. Local (tooth) pain decreased to 50% of original since been seen at NDH • Pain has not gone completely but can be ignored • Ability to sleep • Improvement as reduction in medication and the need to seek medical care at hospital • Relaxation of the jaw • Stability in the bite • Life back: life back under control, back to work, goes out with friends, take her child to the park, back to her home • General relaxation • Ability to cope • Slight but immediate relief of pain after fitting the hard splint • Took ~ 3 m to notice 50-60% reduction in pain	*I feel that now I haven't got TMJD now as such, because that was where the jaw felt damage, it was coming down into my neck, it was giving me earache. I don't have any of those sensations now. The only place I have the pain now is solely in my teeth or the gum areas..* *There are days when my pain is still bad but ... I can put it to the back of my mind to a degree. Not completely, but to a point where you can still get on with your every day life.* *I've managed to go six or seven weeks without having to come up* (to NDH)... *which to me is, a move forward.* *I was pleased to get off the cocktail of medication because I was on so many different things.* *I've got my life back in control. I'm back to work...I go out with my friends... I take my little boy to the park...This sounds really dramatic but Dr Wassell has given me my life back.* *I'm back home. My life's normal. It's back to where it was.* *I can sit back on the couch now put me feet up and watch Eastenders. And I know that in the end the pain's there but I sit on the couch and think I couldn't do this before.* *I'm still in pain really, but I do feel that, you know ...I can live with it.*
Participant 10	• Coping with pain • Pain now is 60% better • Gradual improvement • Slow improvement which stopped (because of the wrong design of the splint) • Feeling happier and better in himself • Sleeping better • Reduction in grinding • It took 6-7m to notice improvement • Then he says it took 3-4w from wearing the splint to notice improvement	*I think I'm coping a lot better than what I was coping with the pain.* *And how long do you think it took to see any improvement? --About six months/seven months.* *I think I felt happier plus I was sleeping better...You know, I wasn't' grinding me teeth as much as I was.*

Perception of care

	Summary of findings	Quotations
Participant 1	Pt was surprised to know at NDH about the link between headaches and grinding Her dentist provided: no treatment, no explanation - just referred her to NDH	*I had headaches, never thinking the headache's from grinding your teeth at first.* *He knew that I needed a lot of treatment and that's when he arranged for me to* (come here).
Participant 2	Received information about her condition only from NDH - no info given by dentist Expectation of Tx. The need for some one to understand and talk to. Treatment provided has not addressed the problem	*I think it was here at the Dental Hospital that they said we'll put you in touch with the Jaw Department in here because it might be something to do with your jaw. It was never me dentist or nothing.* *Just to be cured. Really everybody expects when they go to ... just to cure it like that, but it's not the case is it? It's got to gradually just ...* *I mean you talk to people and they don't understand what you've got. It is a very painful thing, it's not a very nice thing to go through and, I mean, they understand the case when I come... Well nobody really knows what I did. Not a lot of people I've met that ... well what's TMJ? Well what is that? It's obviously something to do with your Jaw so people just sit and look at you. But it is a very painful thing.* *They just kept looking at me teeth, they didn't say anything, just kept look ... drilling into me teeth, injections and drilling into me teeth all the time.* *I'd been to the hospital and I'd been to me own dentist and then here at the dentist and then ... but it's been fine since I've been to the Jaw Clinic... I feel very confident now since I've been here, a lot more confident now.*

Participant 3	First: she thought she is getting older, refused to seek treatment from NDH. Later: was worried / didn't know about it. Thought it is a cancer. Now: she can control the condition Expectations from treatment Dentist has not assured her, did not have a clear idea about it, just referred her, and provided no treatment.	(First): *I thought maybe it's, you know, as you're getting older, a touch of arthritis or anything like that , wear and tear..* (When dentist advised her to go to NDH): *So I said oh it's not that bad I'm not going, you know.* (Later): *I was starting to worry about it, I was imaging everything. Because that's the type of person I am.* (Told her dentist): *"there's something you're not telling me about this"* (After treatment): *I don't think about it much. I really don't. If I get this ache, which is not very often now, I would put that splint in.* (Expectations of treatment): *I don't know what I expected because it had started to improve anyway...I just wanted to know really what it was. I didn't even really think I'd get any sort of treatment for it if that's what it was. I just thought well I'll get my head screened. → I didn't have any x-rays which I assumed I would... And I don't know whether they could feel the actual ... where the jaw joint was, maybe they could feel it moving, or what, but I think that's really about all really...* *The only thing he said "maybe it could be a bit of arthritis" because I had thought that as well, because I have arthritis anyway. He said "but one will only know if you go up the Dental Hospital".* *He mentioned exercises ... but didn't give any exercises*

Participant 4	She thought the symptoms she had were part of the healing process after the teeth removal surgery She was worried of coming back to the hospital when she had a relapse But now she is "ecstatic" No explanations from any one Was aggravated by later dental treatment Referred by her GP to NDH No treatment given by GP or GDP	*And I thought also yes you've had this thing, which is a traumatic thing to have teeth wrenched from your mouth and you're going to have some soreness and pain and the swelling and I thought that that was just part of the jaw healing to be honest. Yet I thought it was taking a long time because even up to about a year afterwards it was quite ... it's actually quite tender under there to touch...I don't know why that happens.. and I thought it was all part of that. And it wasn't until later I thought there's something just, you know ...* *I knew then that there was something that wasn't quite right. I knew I'd had some damage around there, but nothing was ever explained to me about that.* *I think that it may have been aggravated further by treatment I had years later. So I actually had two wisdom teeth removed.*
Participant 5	Dentist avoided responsibility, provided no explanation, no treatment, just referred her to NDH.	*The dentist said that the problem was not to do with the dentist it was something else, you know, she couldn't do anything for me.* *And then I went to my doctor, my GP and he said it was a dental problem and he gave me a slip of paper, wrote something on it and said I should take it to this man and after that she then wrote up here.* *I used to complain to the dentist quite frequently and I was really, really bad and I told her. And she said well it wasn't a problem from her, she said.*
Participant 6	Pt wasn't motivated enough to seek Tx before she got the ear infection Wasn't keen to come to NDH as the problem wasn't serious Dentist provided no explanation, no treatment, just referred her to NDH.	*I was having trouble with my ears. I was having antibiotics for my ears and I just wondered, I said to the Dentist could it come from the jaw. And then he said oh I'll refer you to the hospital, Dental Hospital. I mean I hadn't ... I said that it wasn't serious enough, you know.*

Participant 7	She did not know about her condition before she visited NDH. Assured by the explanation at NDH Dentist provided no explanation, no treatment, just referred her to NDH.	(At the dentist's): *I don't know. They didn't tell me what it was, the TMD.* (At NDH): *They explained about the joints, how the joint worked and that, and some people have just this problem... It's the relief ... because you get it in your mind that it's something ... It's not a worry because I know what it is now.* *My dentist transferred me over here* *they couldn't see what the problem was* *I don't know. They didn't tell me what it was, the TMD, they just thought it was best to come over here and they would look more in depth into it.* *My dentist says it can be quite common and a lot of people are bothered with it like the clicking in the jaw.*
Participant 8	Pt has not informed her dentist as he had told her that her teeth are alright She had no idea about her condition before she visit NDH, and thought she had cancer Assured by the explanation by NDH The dentist made a soft splint (but this was *short and not fitting*) and some occlusal adjustments. He did not provide any explanation about the condition and told her: her teeth are alright	*I must admit I didn't mention it him. I felt as though I was being silly really because he said that my teeth were alright.* *I saw Mr Barclay and he said ... he examined my mouth, obviously, and I had x-rays and so forth and he said what the problem ... there was obviously no real sort of like cancer or anything like that which I was thinking.*

Participant 9	Pt blamed the wrong Tx. Pt thought it was a tumour The need for someone to understand Pt angry about the management before NDH Dentist denied responsibility Pain clinic couldn't help and refer her to the NDH Max fax unit threaten to wire her jaws	*It just seemed to me that it seemed so coincidental that it had come on since the treatment.* *But what happened he'd filled in my tooth over an exposed nerve. He'd drilled too far down... I went to a private dentist and found that out. And the treatment done.* *I didn't know whether I had a tumour or something like that* *My bite's been wrong, which I do believe has been the case all the way through and nobody would believe me.... It's nice to be somewhere where you feel the people can appreciate what you're going through because... like obviously people who, in the dental hospital obviously it's their area of work and it's just a relief to get to somebody... to understand what you're going through.* *So I have had ... I've got a lot of anger still inside me ... about the people in the profession like who didn't believe that they'd done anything wrong, like him who threatened to wire me jaw together when he obviously didn't understand about the bit of the teeth and that me bite was all wrong.* *I went back to my dentist and he said that he'd done all the treatment* *He more or less told us that that wasn't' the case, that it wasn't anything to do with what he'd done, he checked me mouth and said go home, type thing, you must be like grinding your teeth or something.* *This particular registrar, which he was, came in, looked at my mouth, he said well nobody's bite's perfect. He said I think you're just grinding your teeth and you've got acute TMJD. You either stop grinding your teeth or I'm going to have to send you to see a psychologist and even wire your jaw. Well, I mean, when you were in the type of pain that I was in, I just ...* *Dr. X at the pain clinic just put his hands up, he didn't know what to do for my pain.... Nothing.*
Partic. 10	The dentist didn't offer any help at all.	*I went back to the dentist. And she just couldn't ... she couldn't do anything for me.*

Appendix 7
The Oral Health Impact Profile
- adapted version

Over the last month...

1. Have you had difficulty chewing any foods because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

2. Have you had trouble pronouncing any words because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

3. Have you noticed a tooth which doesn't look right?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

4. Have you felt that your appearance has been affected because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

5. Have you felt that your breath has been stale because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

6. Have you felt that your sense of taste has worsened because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

7. Have you had food catching in your teeth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

8. Have you felt that your digestion has worsened because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

Over the last month...

9. Have you felt that your dentures have not been fitting properly?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

10. Have you had painful aching in your mouth?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

11. Have you had a sore jaw?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

12. Have you had headaches because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

13. Have you had sensitive teeth, for example, due to hot or cold foods or drinks?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

14. Have you had toothache?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

15. Have you had painful gums?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

16. Have you found it uncomfortable to eat any foods because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

17. Have you had sore spots in your mouth?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

Over the last month...

18. Have you had uncomfortable dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

19. Have you been worried by jaw or dental problems?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

20. Have you been self conscious because of your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

21. Have jaw or dental problems made you miserable?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

22. Have you felt uncomfortable about the appearance of your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

23. Have you felt tense because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

24. Has your speech been unclear because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

25. Have people misunderstood some of your words because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

26. Have you felt that there has been less flavour in your food because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

Over the last month...

27. Have you been unable to brush your teeth properly because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

28. Have you had to avoid eating some foods because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

29. Has your diet been unsatisfactory because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

30. Have you been unable to eat with your dentures because of problems with them?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

31. Have you avoided smiling because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

32. Have you had to interrupt meals because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

33. Has your sleep been interrupted because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

34. Have you been upset because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

35. Have you found it difficult to relax because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

Over the last month...

36. Have you felt depressed because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

37. Has your concentration been affected because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

38. Have you been a bit embarrassed because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

39. Have you avoided going out because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

40. Have you been less tolerant of your spouse or family because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

41. Have you had trouble getting on with other people because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

42. Have you been a bit irritable with other people because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

43. Have you had difficulty doing your usual jobs because of problems with your teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

Over the last month...

44. Have you felt that your general health has worsened because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

45. Have you suffered any financial loss because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

46. Have you been unable to enjoy other people's company as much because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

47. Have you felt that life in general was less satisfying because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

48. Have you been totally unable to function because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

49. Have you been unable to work to your full capacity because of problems with your jaws, teeth, mouth or dentures?

1	2	3	4	5
Never	**Hardly ever**	**Occasionally**	**Fairly often**	**Very often**

Appendix 8
The Multidimensional Pain Inventory

In the following 20 questions, you will be asked to describe your jaw pain and problems and how it affects your life. Under each question is a scale to record your answer. Read each question carefully and then circle a number on the scale under that question to indicate how that specific question applies to you. An example may help you to better understand how you should proceed to answer these questions.

Example: How frequently do you read the newspaper? (Never) 0 1 2 3 4 5 6 (Daily)
If you never read the newspaper, you would want to circle the number 0. If you read the newspaper nearly every day, you would then circle the number 6. Thus, lower numbers should be used for less frequent newspaper reading, and higher numbers for more frequent reading.

SECTION 1

1. Rate the level of your pain at the present moment.
(No pain) 0 1 2 3 4 5 6 (Very intense pain)

2. In general, how much does your pain or problem interfere with your day to day activities?
(No interference) 0 1 2 3 4 5 6 (Extreme interference)

3. Since the time you developed a pain problem, how much has have your pain or problems changed your ability to work?
(No change) 0 1 2 3 4 5 6 (Extreme change)
Check here, if you have retired for reasons other than your pain problem. ☐

4. How much have your pain or problems changed the amount of satisfaction or enjoyment you get from participating in social and recreational activities?
(No change) 0 1 2 3 4 5 6 (Extreme change)

5. How supportive or helpful is your spouse (significant other) to you in relation to your pain or problems?
(Not at all supportive) 0 1 2 3 4 5 6 (Extremely supportive)

6. Rate your overall mood during the past week.
(Extremely low mood) 0 1 2 3 4 5 6 (Extremely high mood)

7. On the average, how severe have your pain or problems been during the last week?
(No at all severe) 0 1 2 3 4 5 6 (Extremely severe)

8. How much have your pain or problems changed your ability to participate in recreational and other social activities?
(No change) 0 1 2 3 4 5 6 (Extreme change)

9. How much has have your pain or problems changed the amount of satisfaction you get from family-related activities?
(No change) 0 1 2 3 4 5 6 (Extreme change)

10. How worried is your spouse (significant other) about you in relation to your pain problem?

(Not at all worried) 0 1 2 3 4 5 6 (Extremely worried)

11. During the past week how much control do you feel that you have had over your life?

(Not at all in control) 0 1 2 3 4 5 6 (Extremely in control)

12. How much suffering do you experience because of your pain?

(No suffering) 0 1 2 3 4 5 6 (Extreme suffering)

13. How much have your pain changed or problems your marriage and other family relationships?

(No change) 0 1 2 3 4 5 6

14. How much have your pain or problems changed the amount of satisfaction or enjoyment you get from work?

(No change) 0 1 2 3 4 5 6 (Extreme change)

Check here, if you are not presently working ☐.

15. How attentive is your spouse (significant other) to your pain problem?

(Not at all attentive) 0 1 2 3 4 5 6 (Extremely attentive)

16. During the past week how much do you feel that you've been able to deal with your problems?

(Not at all) 0 1 2 3 4 5 6 (Extremely well)

17. How much have your pain or problems changed your ability to do household chores?

(No change) 0 1 2 3 4 5 6 (Extreme change)

18. During the past week how irritable have you been?

(Not at all irritable) 0 1 2 3 4 5 6 (Extremely irritable)

19. How much have your pain or problems changed your friendships with people other than your family?

(No change) 0 1 2 3 4 5 6 (Extreme change)

20. During the past week how tense or anxious have you been?

(Not at all tense or anxious) 0 1 2 3 4 5 6 (extremely tense and anxious)

SECTION 2
In this section, we are interested in knowing how your spouse (or significant other) responds to you when he or she knows that you are in pain.
On the scale listed below each question, circle a number to indicate how often your spouse for significant other) responds to you in that particular way when you are in pain. Please answer all of the 14 questions.

1. Ignores me.
(Never) 0 1 2 3 4 5 6 (Very often)

2. Asks me what he/she can do to help.
(Never) 0 1 2 3 4 5 6 (Very often)

3. Reads to me.
(Never) 0 1 2 3 4 5 6 (Very often)

4. Expresses irritation at me.
(Never) 0 1 2 3 4 5 6 (Very often)

5. Takes over my jobs or duties.
(Never) 0 1 2 3 4 5 6 (Very often)

6. Talks to me about something else to take my mind off the pain.
(Never) 0 1 2 3 4 5 6 (Very often)

7. Expresses frustration at me.
(Never) 0 1 2 3 4 5 6 (Very often)

8. Tries to get me to rest.
(Never) 0 1 2 3 4 5 6 (Very often)

9. Tries to involve me in some activity.
(Never) 0 1 2 3 4 5 6 (Very often)

10. Expresses anger at me.
(Never) 0 1 2 3 4 5 6 (Very often)

11. Gets me some pain medications.
(Never) 0 1 2 3 4 5 6 (Very often)

12. Encourages me to work on a hobby.
(Never) 0 1 2 3 4 5 6 (Very often)

13. Gets me something to eat or drink.
(Never) 0 1 2 3 4 5 6 (Very often)

14. Turns on the T.V. to take my mind off my pain.
(Never) 0 1 2 3 4 5 6 (Very often)

SECTION 3
Listed below are 18 common daily activities. Please indicate how often you do each of these activities by circling a number on the scale listed below each activity. Please complete all 18 questions.

1. Wash dishes.
(Never) 0 1 2 3 4 5 6 (Very often)

2. Mow the lawn.
(Never) 0 1 2 3 4 5 6 (Very often)

3. Go out to eat.
(Never) 0 1 2 3 4 5 6 (Very often)

4. Play cards or other games.
(Never) 0 1 2 3 4 5 6 (Very often)

5. Go grocery shopping.
(Never) 0 1 2 3 4 5 6 (Very often)

6. Work in the garden.
(Never) 0 1 2 3 4 5 6 (Very often)

7. Go to a movie.
(Never) 0 1 2 3 4 5 6 (Very often)

8. Visit friends.
(Never) 0 1 2 3 4 5 6 (Very often)

9. Help with the house cleaning.
(Never) 0 1 2 3 4 5 6 (Very often)

10. Work on the car.
(Never) 0 1 2 3 4 5 6 (Very often)

11. Take a ride in a car.
(Never) 0 1 2 3 4 5 6 (Very often)

12. Visit relatives.
(Never) 0 1 2 3 4 5 6 (Very often)

13. Prepare a meal.
(Never) 0 1 2 3 4 5 6 (Very often)

14. Wash the car.
(Never) 0 1 2 3 4 5 6 (Very often)

15. Take a trip.
(Never) 0 1 2 3 4 5 6 (Very often)

16. Go to a park or beach.
(Never) 0 1 2 3 4 5 6 (Very often)

17. Do a load of laundry.
(Never) 0 1 2 3 4 5 6 (Very often)

18. Work on a needed house repair.
(Never) 0 1 2 3 4 5 6 (Very often)

Appendix 9
Participant information sheet and consent form used in *Study Three*

Patients' copy

UNIVERSITY OF
NEWCASTLE UPON TYNE

Participants Information Sheet

Development of a Quality of Life Instrument (QOL) To Assess Treatment Outcome of Temporomandibular Disorders (TMD)

How do your jaw problems impact on everyday life?

How do things change over time and with treatment?

You are being invited to take part in a research study, aimed to find out how your symptoms influence your life and how changes occur over time and with treatment. Before you decide to participate it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What is the purpose of the study?
Temporomandibular disorders (TMD) are a set of uncomfortable and painful symptoms either in the jaw muscles, jaw joints or both. The aim of this study is to develop a questionnaire to measure the impact of these jaw problems on patients' quality of life before and after treatment.

Why have I been chosen for the study?
You have been diagnosed as having TMD jaw problems. You are, thus, invited to take part in this study. We are interested in investigating how your life is affected by TMD, and monitor changes occurring with treatment. Altogether 100 patients and 100 TMD free individuals will be invited to participate.

What will happen to me if I take part and what do I have to do?
You are asked to complete a package of three questionnaires that evaluate your pain, discomfort and how your jaw problem impacts on your everyday life. The questionnaires will be given twice, now and 4 months later. On the second occasion, we may send you the questionnaires by post or ask you to fill them in the next time you attend this clinic. The questionnaires can usually be answered within 30 minutes.

Your records in the dental hospital will be examined by the researcher to confirm the diagnosis and to monitor your treatment progress. It is important to highlight that the study has no influence on the treatment you are receiving in any way.

What are the possible benefits of taking part?
The information gained from this study will be used to develop a method of measuring how TMD patients respond to different treatments. This will be an important contribution in determining which are the best treatments. We cannot guarantee that your participation in this study will be of direct benefit to you, and we are unable to pay a fee for your participation in this study.

Do I have to take part?
It is up to you to decide whether or not to take part in this study. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive in any way.

Will my taking part in this study be kept confidential?
Your name will not be written on any of the questionnaires, but a code number; so that we have no direct way of knowing which package belongs to which person.
Personal details and opinions you may express will be kept strictly confidential in accordance with the provisions of the Data Protection Act (1998). Your answers will be stored in a locked cupboard with access only available to the researcher who has asked you to participate.
Afterwards, the questionnaires will be analysed, and the results of this study may be published in a scientific journal or presented at a research conference. In either case, your name will not be mentioned as part of the publication.

Who is organising and funding the research?
Who has reviewed the study?
This study has been organised by the School of Dental Sciences - Newcastle University, and is being carried out as part of a clinical PhD programme. It has been reviewed by Newcastle & North Tyneside Local Research Ethics Committee1.

Contact for Further Information
If you have any questions please don't hesitate to ask either the researcher* or the consultant in charge of the clinic**.

We would like finally, to thank you for the time you spent reading this sheet, and, should you agree for participating in our study. It is much appreciated.

Version 2
Date: April 2004

---

* Dr Adel MOUFTI. Dept. of Restorative Dentistry, The School of Dental Sciences, University of Newcastle upon Tyne, NE2 4BW
** Dr RW Wassell and Prof. JG Steele, Senior Lecturers & Honorary Consultants. Dept. of Restorative Dentistry, The School of Dental Sciences, University of Newcastle upon Tyne, NE2 4BW
Tel 0191 222 7825 - Fax 0191 222 6137

Controls' copy

UNIVERSITY OF
NEWCASTLE UPON TYNE

Participants Information Sheet

Development of a Quality of Life Instrument (QOL) To Assess Treatment Outcome of Temporomandibular Disorders (TMD)

How do jaw problems impact on everyday life?

You are being invited to take part in a research study, aimed to find out how jaw problems influence patients' everyday life. **At the outset we should explain that we are looking for control subjects who do not have these problems so that we can establish a baseline for comparison.**

Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What is the purpose of the study?
Temporomandibular disorders (TMD) are a set of uncomfortable and painful symptoms either in the jaw muscles, jaw joints or both. The aim of this study is to develop a questionnaire to measure the impact of these jaw problems on patient's quality of life before and after treatment.

Why have I been chosen for the study?
The researcher has asked you a couple of questions and found that you do not have a jaw problem. You are, thus, invited to take part in this study as a control subject. Your participation will help us exploring the differences between TMD patients and people like you who do not suffer with jaw problems. Altogether 100 patients and 100 control subjects will be invited to participate.

What will happen to me if I take part and what do I have to do?
You are asked to answer two questionnaires that evaluate dental & oral pain and discomfort, and the impact of those on everyday life. **The questionnaires contain items that apply to every one including TMD patients and control subjects who may have other problems**

with their teeth and mouth. You will need to answer the questions as accurately as possible to enable us find out how you feel. In total, the questions can be answered within 30 minutes.

What are the possible benefits of taking part?
The information gained from this study will be used to develop a method of measuring how TMD patients respond to different treatments. This will be an important contribution in determining which are the best treatments. We cannot guarantee that your participation in this study will be of direct benefit to you, and we are unable to pay a fee for your participation in this study.

Do I have to take part?
It is up to you to decide whether or not to take part in this study. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect access to treatment or standard of care in any way.

Will my taking part in this study be kept confidential?
Your name will not be written on any of the questionnaires, but a code number; so that we have no direct way of knowing which answer sheet belongs to which person.
Personal details and opinions you may express will be kept strictly confidential in accordance with the provisions of the Data Protection Act (1998). Your answers will be stored in a locked cupboard with access only available to the researcher who has asked you to participate.
Afterwards, the questionnaires will be analysed, and the results of this study may be published in a scientific journal or presented at a research conference. In either case, your name will not be mentioned as part of the publication.

Who is organising and funding the research?
Who has reviewed the study?
This study has been organised by the School of Dental Sciences - Newcastle University, and is being carried out as part of a clinical PhD programme. It has been reviewed by Newcastle & North Tyneside Local Research Ethics Committee1.

Contact for Further Information
If you have any questions please don't hesitate to ask either the researcher* or the consultant in charge of the clinic**.

We would like finally, to thank you for the time you spent reading this sheet, and, should you agree, for participating in our study. It is much appreciated.

Version 2
Date: April 2004

---

* Dr Adel MOUFTI. Dept. of Restorative Dentistry, The School of Dental Sciences, University of Newcastle upon Tyne, NE2 4BW
** Dr RW Wassell and Prof. JG Steele, Senior Lecturers & Honorary Consultants. Dept. of Restorative Dentistry, The School of Dental Sciences, University of Newcastle upon Tyne, NE2 4BW
Tel 0191 222 7825 - Fax 0191 222 6137

Participant identification number for this study: .....
Ethics reference: 04/Q0905/24

CONSENT FORM

Development of a Quality of Life Instrument (QOL) To Assess Treatment Outcomes of Temporomandibular Disorders (TMD)

Name of Researcher: Mr. Adel Moufti

Please initial box

1. I confirm that I have read and understand the information sheet Version 2 dated April 2004 for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

(For TMD patients only): ☐

3. I understand that sections of any of my medical notes may be looked at by responsible individuals from the Dental School in Newcastle University or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records.

4. I agree to take part in the above study ☐

Name of Participant	Date	Signature
Researcher	Date	Signature

1 for patient; 1 for researcher; 1 to be kept with hospital notes (for patients only).

Appendix 10
Form used to check inclusion criteria for controls in *Study Three*

University of Newcastle – School of Dental Sciences 2004 *Confidential*

Participant's code number 0☐☐☐ **Date** ☐☐ ☐☐ 2☐0☐

Age ☐☐ **Sex** ☐

[To be completed by the researcher]

1- Do you have any pain or problems with your jaw muscles or joints for which you need treatment?

Circle one choice:

No I do not need treatment → Proceed to next question

Yes I need treatment → Not suitable as a control

2- Are you under treatment for pain or problems with your jaw muscles or joints?

Not under treatment → Proceed to next page

Under treatment: **Currently** → Not suitable as a control

Recently → Not suitable as a control