

**You will never find time for anything.
If you want time, you must make it.**

Charles Buxton (1823- 1871)

The replacement of plastic restorations in clinical dental practice and the effect that a simple training programme can have on the decision making process

**Robert McAndrew
BDS (Dund), MscD (Wales), FDS,
FDS(Rest Dent)RCS, DRD, MRDRCS(Ed)**

A thesis presented in fulfilment for the degree of Doctor in Philosophy,
University of Cardiff, 2008.

UMI Number: U584353

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U584353

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

I would like to thank and acknowledge the assistance from everyone who has helped me in some way, however small that may have been, towards the final production of this work; unfortunately this is not practical and I would only forget someone. So to all of you whom I do not mention below, I apologise unreservedly.

My deepest gratitude and heartfelt thanks go to Professor Barbara Chadwick and Professor Elizabeth Treasure for, without doubt, their guidance, advice and inspiration have been an integral part in this submission. To continue to learn from their words of wisdom and sound advice will be testimony to them; I hope I do so in the future

To the “gang of 21” and the volunteer patients, I offer my deepest gratitude. I will be forever indebted to their commitment and willingness to help me out. A special thank goes to Chris Fenton for helping with the model production.

To friends, colleagues and family (somewhere between 53 and 68 at the last count!) and the makers of Titleist™ Pro V1 golf balls...thanks for putting up with what occasionally must have felt like the ranting of a lunatic.

DEDICATION

I would like to dedicate this thesis to all those who have encouraged me in my academic and clinical careers over the years; to those who ever dared to doubt...pah!

TABLE OF CONTENTS

Quotation.....	i
Title page.....	ii
Acknowledgements.....	iii
Declaration.....	iv
Dedication.....	v
Table of contents.....	vi
List of tables.....	xii
List of figures.....	xv
Abstract.....	xx

CHAPTER 1 INTRODUCTION.....	1
Section 1.1 Introduction.....	2

CHAPTER 2 REVIEW OF THE LITERATURE.....	5
--	----------

Section 2.1 Presenting the challenges.....	6
2.1.1 Defining failure.....	6
2.1.2 The reasons for restoration replacement.....	7
2.1.3 The prevalence of restoration replacement.....	9
2.1.4 Estimating how long a restoration will last.....	9
2.1.5 The cost of restoration replacement.....	11
2.1.6 Variation in decision making.....	12
2.1.7 Trying to determine failure.....	12
2.1.8 Defining, determining and assessing quality of dental restorations.....	15
<i>Defining and determining quality</i>	15
<i>Assessing quality</i>	16
2.1.9 The USPHS criteria.....	18
<i>Reasons for using USPHS</i>	18
<i>Reasons against using USPHS</i>	20
<i>Using USPHS</i>	23
<i>Assessing caries</i>	24
<i>Assessing marginal discolouration</i>	25
<i>Assessing marginal integrity</i>	26
<i>Assessing anatomical form</i>	27
<i>Assessing aesthetics</i>	27
2.1.10 Calibration in clinical research.....	28
2.1.11 Summary.....	30

Section 2.2 How long does a restoration last? An evaluation of the literature by examining study design.....	31
2.2.1 Cross sectional studies and the challenges they pose.....	34
<i>Challenges relating to sampling and response rates</i>	35
<i>Challenges relating to subjective data analysis</i>	36

	<i>Challenges relating to materials and /or clinical procedure</i>	37
	<i>Challenges relating to the patient's past dental experience</i>	38
	<i>Challenges relating to the time of sampling</i>	38
	<i>Challenges relating to data manipulation</i>	38
	<i>The value of cross sectional studies in evaluating restoration longevity</i>	40
	<i>Summary on cross sectional studies</i>	40
2.2.2	<i>Cohort studies and the challenges they pose</i>	42
2.2.2a	<i>Prospective cohort studies</i>	43
	<i>Challenges relating to operator numbers</i>	43
	<i>Challenges relating to patient selection</i>	44
	<i>Challenges relating to changes in dental materials</i>	44
	<i>Challenges relating to clinical technique</i>	45
	<i>Challenges relating to tooth variables</i>	45
	<i>Challenge relating to drop-out</i>	45
	<i>Challenges relating to data manipulation</i>	46
	<i>Challenges relating to data comparison</i>	47
2.2.2b	<i>Retrospective cohort studies</i>	48
	<i>Challenges relating to data collection</i>	48
	<i>Summary on cohort studies</i>	50
2.2.3	<i>Clinical trials and the challenges they pose</i>	52
	<i>The challenges posed by effect modifiers</i>	52
	<i>The challenge of bias</i>	53
	<i>The challenge of blinding</i>	53
	<i>The challenge of randomisation</i>	54
	<i>Challenges associated with the operator</i>	55
	<i>The challenge of environment</i>	55
	<i>The challenge posed by multiple subgroups</i>	57
	<i>Challenges relating to operative procedure and material variability</i>	58
	<i>The challenge associated with standardising measuring and recording</i>	59
	<i>The challenge of obtaining standard cavity types</i>	61
	<i>The challenge of follow up</i>	62
	<i>Challenges related to identifying and measuring failure</i>	63
	<i>The challenge of time</i>	64
	<i>Summary on clinical trials</i>	65
2.2.4	<i>Problems associated with different types of clinical study: An evaluation of the challenges by examining the literature by limiting the field</i>	67
	<i>Prospective case series using restoration replacement criteria without training</i>	70
	<i>Prospective case series using restoration replacement using replacement criteria with training</i>	70
	<i>Prospective case series using restoration replacement using validated outcome criteria</i>	71
	<i>Prospective case series using restoration replacement and measuring "true" restoration failure</i>	72
	<i>Retrospective study with concurrent controls using replacement criteria without training</i>	73

<i>Prospective study with historical controls using replacement criteria without training</i>	74
<i>Prospective study with historical controls and measuring "true" restoration failure</i>	74
<i>Prospective study with concurrent controls using replacement criteria without training</i>	75
<i>Prospective study with concurrent controls using replacement criteria with training</i>	76
<i>Prospective study with concurrent controls using validated outcome criteria</i>	77
<i>Prospective study with concurrent controls and measuring "true" restoration failure</i>	77
<i>Clinical trial using replacement criteria without training</i>	78
<i>Clinical trial using replacement criteria with training</i>	78
<i>Clinical trial using validated outcome criteria</i>	79
<i>Clinical trial measuring "true" restoration failure</i>	80
<i>Randomised controlled clinical trial using replacement criteria without training</i>	81
<i>Randomised controlled clinical trial using replacement criteria with training</i>	81
<i>Randomised controlled clinical trial using validated outcome criteria</i>	82
<i>Randomised controlled clinical trial measuring true" restoration failure</i>	83
<i>Exclusions</i>	83
<i>Justifications for not including papers with subjective opinion as the determinant of outcome</i>	85
<i>Justifications for not including papers based on case study's in a systematic review</i>	85

Section 2.3 Outlining the research	85
---	----

Section 2.4 Tables for chapter 2	88
---	----

CHAPTER 3 MATERIALS AND METHOD	95
---------------------------------------	----

Section 3.1 Overall aim	96
--------------------------------	----

Section 3.2 Preparatory phase	97
--------------------------------------	----

3.2a Selection of teeth and manufacture of models	97
---	----

3.2b Determining and setting the gold standard for restoration replacement	99
--	----

3.2c The recruitment and selection of dentists to take part in the study	100
--	-----

3.2d The design and manufacture of training material	101
--	-----

Section 3.3 The simulated-clinical phase	102
---	-----

3.3a The assessment of the restorations by the untrained dentists	102
---	-----

3.3b The selection and training of the dentists	104
---	-----

3.3b.i The training programme	104
-------------------------------	-----

3.3b.ii Evaluating the effects training	105
---	-----

Section 3.4 The pre-clinical phase: including the recruitment of patients and the determination of the gold standard for the clinical phase	106
Section 3.5 The clinical phase: including the assessment of the restorations recruited into the clinical phase by the trained and untrained dentists	107
Section 3.6 Post-clinical phase	108
3.6a Confirmation of the gold standard and determination of no harm to the recruited patients	108
3.6b An evaluation of procedure within the study by the participants	109
Section 3.7 Statistical considerations	109
Section 3.8 Tables for chapter 3	114
Section 3.9 Figures for chapter 3	119

CHAPTER 4 RESULTS

Section 4.1 Determining the gold standard for restoration replacement: Assessments recorded and statistical analysis	130
Section 4.2 Basic data collected on the group participating in the study	132
Section 4.3 Simulated clinical phase	134
4.3a Pre-training simulated clinical phase: Assessments recorded and statistical analysis	134
4.3b Pre-training simulated clinical phase: assessments recorded and statistical analysis but separated after allocation to training or control groups	137
4.3c Post-training simulated clinical phase: results after training the test group	138
Section 4.4 Clinical phase: Assessments recorded and statistical analysis	141
Section 4.5 Post participation evaluation	143
Section 4.6 Tables for chapter 4	145
Section 4.7 Figures for chapter 4	180

CHAPTER 5 DISCUSSION

Section 5.1 Pre-simulated clinical phase	184
5.1a Selection of teeth and construction of models	184
5.1b Determining the gold standard for restoration replacement	190
5.1c The recruitment and selection of dentists to take part in the study	195
Section 5.2 Simulated clinical phases	198

5.2a	The assessment of the restorations by the untrained dentists.....	198
5.2b	The selection and training of the dentists.....	200
5.2c	Manufacture of training material.....	201
5.2d	The training programme.....	202
5.2e	Evaluating the effects of training in the simulated clinical phase.....	204
Section 5.3	The clinical phase.....	209
5.3a	Recruitment of patients and determination of the gold standard for the clinical phase.....	209
5.3b	Confirmation of gold standard and determination of no harm.....	215
Section 5.4	An evaluation of procedure within the study by the participants.....	216
Section 5.5	General discussion point: Utilising USPHS as a clinical tool.....	219
Section 5.6	Tables for chapter 5.....	222
Section 5.7	Figures for chapter 5.....	224

CHAPTER 6 SUMMARY, CONCLUSIONS AND FURTHER STUDY.....	231
--	------------

Section 6.1	Conclusions.....	232
Section 6.2	Further study.....	235

CHAPTER 7 BIBLIOGRAPHY.....	237
------------------------------------	------------

APPENDIX		xxi
Appendix 1	The USPHS criteria	xxii
Appendix 2	Examples of results sheets and graphic material used	xxiii
2.1	Result sheet for determining the “gold standard”	xxxix
2.2	Initial assessment sheet for assessors	xxx
2.3	Graphic plates used in training	
2.3a	Caries Assessment	xxxi
2.3b	Marginal Adaptation	xxxii
2.3c	Marginal Discolouration	xxxiii
2.3d	Colour Match	xxxiv
2.3e	Anatomical Form	xxxv
2.4	Examples of photographs used in the training programme	xxxvi
Appendix 3	Consent forms and copies of approval requests	xxxvii
3.1a	Ethical approval	xxxviii
3.1b	Trust utilisation of resource approval	xxxix

3.2a	Consent sheet for volunteering to take part in the project.....	xi
3.2b	Consent form for taking part in phase 1	xli
3.2c	Consent sheet for volunteer patients.....	xlii
3.2d	Patient information sheet.....	xliii
3.3	Photographic consent form.....	xliv
Appendix 4	An explanation of calculating Dices Coincidence Index	xlvi
Appendix 5	Data used to calculate average experience of registered dentists in the United Kingdom.....	xlvi

LIST OF TABLES

CHAPTER 2 LITERATURE REVIEW

Table 2.1	Reasons stated in the literature for replacing restorations	89
Table 2.2	An indication of the similarities in restoration replacement rates between different countries over the decades.....	90
Table 2.3	Format of restoration failure (Kreulen <i>et al</i> , 1998).....	91
Table 2.4	Cross sectional studies reporting data on the longevity of amalgam restorations.....	92
Table 2.5	Cohort studies looking at the longevity of amalgam restorations.....	93
Table 2.6	Papers accepted into the systematic review by Chadwick <i>et al</i> (2001).....	94

CHAPTER 3 MATERIAL AND METHODS

Table 3.1	Statistical parameters calculated in the phases of the study	115
Table 3.2	Kappa values and related estimates of strength of agreement (Landis and Koch, 1977).....	116
Table 3.3	A 2x2 contingency table illustrating how positive and negative predictive values are calculated.....	117

Table 3.4	Explanation of how Dice's Coincidence Index is calculated	118
------------------	---	-----

CHAPTER 4 RESULTS

Table 4.1.1a	1 st Assessment of restorations used for the simulated clinical phase (BC)	146
Table 4.1.1b	1 st Assessment of restorations used for the simulated clinical phase (RM)	147
Table 4.1.1c	2 nd Assessment of restorations used for the simulated clinical phase (BC)	148
Table 4.1.1d	2 nd Assessment of restorations used for the simulated clinical phase (RM)	149
Table 4.1.1e	Agreement and disagreements between the assessments of the restorations made by the gold standard assessors during the 1 st assessment	150
Table 4.1.1f	Agreement and disagreements between the assessments of the restorations made by the gold standard assessors during the 2 nd assessment	151
Table 4.1.2	Summary data for restoration replacements decisions made by the gold standard assessors on the restorations used in the simulated clinical phases of the study	152
Table 4.1.3	Results achieved by the gold standard assessors when compared to themselves, each other and the agreed gold standard	153
Table 4.1.4	Results achieved by the gold standard assessors when compared to themselves and the agreed gold standard for restoration replacement	153
Table 4.2.1	Basic demographic details for the randomly selected volunteers	154
Table 4.3	Raw data resulting from the initial examination of the restorations in the simulated clinical phase prior to any training phase	155
Table 4.3.1	Pre-training simulated clinical phase results evaluated by demographic parameters and expressed as mean, standard deviation and significance (Mann Whitney U test)	156

Table 4.3.2	Pre-training simulated clinical phase results evaluated by demographic parameters and expressed as mean, standard deviation and significance (Mann Whitney U test).....	157
Table 4.3.3	Pre-training simulated clinical phase results evaluated by demographic parameters and expressed as mean, standard deviation and significance (Mann Whitney U test).....	158
Table 4.3.4	Differences between the trained (test) and untrained (control) group during the pre-training simulated clinical phase of the study.....	159
Table 4.4.1	Raw data for the assessors showing the differences for the test (trained) group before and after training plus the control (untrained) group during the simulated clinical phases of the study.....	160
Table 4.4.2	Comparisons of the washout periods between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study.....	161
Table 4.4.3	Comparisons of the full examination times taken between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study.....	162
Table 4.4.4	Comparisons of the time taken for the intra-examiner kappa statistic examinations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study.....	163
Table 4.4.5	Comparisons of the number of restorations scheduled for replacement between the untrained (control) group, untrained (test) group and the trained (test) group during the laboratory phases of the study	164
Table 4.4.6	Comparison of the Dice's Coincidence Index for agreements that a restoration was sound between the untrained (control) group, untrained (test) group and the trained (test) group during the laboratory phases of the study.....	165
Table 4.4.7	Comparison of the Dice's Coincidence Index for agreements that a restoration required replacement between the untrained (control) group, untrained (test) group and the trained (test) group during the laboratory phases of the study.....	166
Table 4.4.8	Comparisons of the sensitivity calculations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases	

of the study	167
Table 4.4.9 Comparison of the specificity calculations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study	168
Table 4.4.10 Comparison of the positive predictive value calculations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study	169
Table 4.4.11 Comparison of the negative predictive value calculations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study	170
Table 4.4.12 Comparison of the intra-examiner kappa statistic for variability between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study	171
Table 4.4.13 Comparison of the inter-examiner kappa statistic for variability between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study	172
Table 4.5.1 Summary data for patients taking part in the clinical phase	173
Table 4.5.2 Raw data, evaluations and subsequent statistical calculations for the clinical phase	174
Table 4.5.3 Means, standard deviations and results of the statistical comparisons of the means (Mann-Whitney U test) for the untrained (control) group and the trained (test) group for the clinical phase of the project	175
Table 4.6.1 Responses to the evaluation survey	178
Table 4.6.2 Free comments generated by the evaluation questionnaire	179

CHAPTER 5 DISCUSSION

Table 5.1 Research supporting USPHS use as a valuable clinical tool (adapted from Chadwick <i>et al</i> , 2001)	223
--	-----

LIST OF FIGURES

CHAPTER 3 MATERIAL AND METHODS

Figure 3.1	Summary of materials and method	120
Figure 3.2	Timeline for study	121
Figure 3.3	An example of replica jaws used in the study	122
Figure 3.4	Type and distribution of restorations used for the laboratory phases	123
Figure 3.5	Replica jaws in phantom head	124
Figure 3.6	Phantom head with replica jaws and teeth being used in the clinic	125
Figure 3.7	Type and distribution of restorations in patients used for the clinical phase	126
Figure 3.8	Evaluation proforma	127
Figure 3.9	Q-Q plot showing distribution of data	128

CHAPTER 4 RESULTS

Figure 4.1a	Mean number of restorations scheduled for replacement during the pre-training simulated clinical phase	156
Figure 4.1b	Time taken to complete the initial examination during the pre-training simulated clinical phase	156
Figure 4.1c	Time taken to evaluate the restorations for the “kappa” determinations during the pre-training simulated clinical phase	156
Figure 4.1d	Time between the initial and re-examinations during the pre-training simulated clinical phase	156

Figure 4.2a	Mean Dice's coincidence index values showing the degree of agreement between examiners that a restoration did or did not require replacement during the pre-training simulated clinical phase.....	157
Figure 4.2b	Mean sensitivity and specificity values for the examiners during the pre-training simulated clinical phase.....	157
Figure 4.3a	Mean positive and negative predictive values illustrating the agreement between examiners during the pre-training simulated clinical phase.....	158
Figure 4.3b	Mean kappa agreement calculations for the examiners during the pre-training simulated clinical phase showing the agreement with themselves and the gold standard.....	158
Figure 4.4.2a	Box plots showing the differences in washout period between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	161
Figure 4.2.2b	Box plots showing the differences in washout period between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	161
Figure 4.4.3a	Box plots showing the differences in full examination times between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	162
Figure 4.4.3b	Box plots showing the differences in full examination times between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	162
Figure 4.4.4a	Box plots showing the differences in time taken for the intra-examiner kappa statistic examinations between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	163
Figure 4.4.4b	Box plots showing the differences in time taken for the intra-examiner kappa statistic examinations between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	163
Figure 4.4.5a	Box plots showing the differences in number of restorations scheduled for replacement between the untrained (control) group and the untrained (test) group for the simulated clinical phase of the study.....	164

Figure 4.4.5b	Box plots showing the differences in number of restorations scheduled for replacement between the untrained (test) group and the trained (test) group for the simulated clinical phase of the study.....	164
Figure 4.4.6a	Box plots showing the differences in full examination times between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	165
Figure 4.4.6b	Box plots showing the differences in full examination times between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	165
Figure 4.4.7a	Box plots showing the differences in Dice's Coincidence Index for agreements that a restoration required replacement between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	166
Figure 4.4.7b	Box plots showing the differences in Dice's Coincidence Index for agreements that a restoration required replacement between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	166
Figure 4.4.8a	Box plots showing the differences in sensitivity between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	167
Figure 4.4.8b	Box plots showing the differences in sensitivity between the untrained (test) group and the untrained (test) group for the simulated clinical phase.....	167
Figure 4.4.9a	Box plots showing the differences in specificity between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	168
Figure 4.4.9b	Box plots showing the differences in specificity between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	168
Figure 4.4.10a	Box plots showing the differences in positive predictive values between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	169
Figure 4.4.10b	Box plots showing the differences in positive predictive values between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	169
Figure 4.4.11a	Box plots showing the differences in negative predictive values between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	170

Figure 4.4.11b	Box plots showing the differences in negative predictive values between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	170
Figure 4.4.12a	Box plots showing the differences in intra-examiner kappa statistic between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	171
Figure 4.4.12b	Box plots showing the differences in intra-examiner kappa statistic between the trained (test) group and the untrained (test) group for the simulated clinical phase.....	171
Figure 4.4.13a	Box plots showing the differences in inter-examiner kappa statistic between the untrained (control) group and the untrained (test) group for the simulated clinical phase.....	172
Figure 4.4.13b	Box plots showing the differences in inter-examiner kappa statistic between the untrained (test) group and the trained (test) group for the simulated clinical phase.....	172
Figure 4.5.3a	A box plot showing the difference between the untrained control) group and the trained (test) group with respect to the number of restorations scheduled for replacement in the clinical phase of the project.....	176
Figure 4.5.3b	A box plot showing the difference between the untrained (control) group and the trained (test) group with respect to agreeing with the gold standard set for the clinical phase of the project.....	176
Figure 4.5.3c	A box plot showing the difference between the untrained (control) group and the trained (test) group with respect to the time taken to examine the full cohort of patients during the clinical phase of the project.....	177
Figure 4.5.3d	A box plot showing the difference between the untrained (control) group and the trained (test) group with respect to the time taken to examine patients during the intra-examiner kappa statistic examinations of patients in the clinical phase of the project.....	177
Figure 4.1	Examples of restorations where the gold standard operators disagreed.....	181
Figure 4.2	Graph showing the number of assessors by the number of years qualified.....	182

CHAPTER 5 DISCUSSION

- Figure 5.1** Graphical representation of Table 3.1.2c showing the agreements and disagreements between the gold standard assessors 225
- Figure 5.2** Restoration 20793L6O, the only restoration not to be scheduled for replacement by both the gold standard assessors on the second examination 226
- Figure 5.3** Examples of restorations where none of the untrained assessors disagreed with the gold standard that the restorations required replacement 227
- Figure 5.4** Examples of restorations where more than five of the untrained assessors disagreed with the gold standard for restoration replacement 228
- Figure 5.5** Examples of restorations where none of the untrained assessors disagreed with the gold standard that the restorations did not need to be replaced 229
- Figure 5.6** Examples of restorations where the untrained assessors' disagreed with the gold standard that the restorations did not need replaced occurred 239

ABSTRACT

AIM: To investigate how a simple training programme affected restoration replacement decision making by a group of sixteen dentists. **METHOD:** This project had two distinct phases, one involving a simulated clinical examination of 111 restorations and the other a clinical examination of 66 restorations. On both occasions, two experienced clinicians using the United States Public Health Service (USPHS) criteria determined the restorative status of the restorations; these evaluations determined the gold standard status with respect to restoration integrity for the restorations. All evaluations were completed under strictly controlled clinical conditions with standard equipment and lighting provided. After completing the simulated clinical phase half of the sixteen dentists were randomly assigned to undertake restoration evaluation training (test group). The results of the simulated clinical and clinical examinations between the test and the non-trained group (control group) were compared by the non-parametric statistical analysis of a number of parameters i.e. the number of restorations scheduled for replacement, the time taken to complete examinations, sensitivity, specificity, positive predictive value, negative predictive value, Dice's coincidence index and Cohen's Kappa statistic. **RESULTS:** There were no statistical differences between the groups at baseline with test and control groups scheduling a similar number of restorations for replacement (36.25 ± 7.78 and 34.75 ± 7.93). After training the test group took longer to complete a repeat simulated clinical examination; 59.25 ± 5.06 minutes, when this was compared with that of the initial examination, 39.13 ± 8.54 minutes. However, there were no other statistically significant differences when baseline measurements were compared. The clinical phase highlighted a number of statistically significant results when the test and the control group were compared; the number of restorations scheduled for replacement (6.00 ± 3.01 and 9.71 ± 3.15), examination time (27.86 ± 3.45 and 36.71 ± 3.74) and agreement with the gold standard for restoration replacement (0.85 ± 0.27 and 0.79 ± 0.06). **CONCLUSION:** Within the limits of this study, it was concluded that examiner training can have a significant effect on plastic restoration replacement decision making by dentists.

CHAPTER 1

INTRODUCTION

1. Introduction

It is well accepted that **all** dental restorations **will eventually** fail and that this occurs despite technical excellence (Jokstad *et al*, 2001). Over half of dental restorations placed in the United Kingdom are replacements (Clarkson *et al*, 2000; Burke *et al*, 1999; Wilson *et al*, 1997a; Nuttall and Elderton, 1983) and similar figures have been reported over the decades and throughout the world (Table 2.1)

The literature highlights that the failure of dental restorations is of major concern in dental practice (Manhart *et al*, 2004) and there are many clinical studies that detail why restorations are replaced (Table 2.2) The performance of dental restorations relates to a number of factors; the restorative material (Mair, 1998; Mjör and Jokstad, 1993), clinician's experience (Mjör *et al*, 2000), tooth type and position in the arch (Kolker *et al*, 2004; Johnson *et al*, 1992; Norman *et al*, 1990; Drake, 1988), size and design of restoration (Lucarotti *et al*, 2005b; Wahl *et al*, 2004; Jokstad and Mjör, 1991b; Norman *et al*, 1990) and patient age (Wahl *et al*, 2004; Mjör *et al*, 2000). Previously, Elderton (1976a) listed 22 reasons for restoration failure but generally they fit into one of three groups; material failure, operator error and patient factors (Manhart *et al*, 2004).

Many aspects of restorative care are subject to variation within and between examiners; for example, treatment planning (Grembowski *et al*, 1991; Kay *et al*, 1988; Elderton and Nuttall, 1983; Bailit and Clive, 1981; Bailit *et al*, 1979; Rytömaa *et al*, 1979), periodontal assessment (Walsh and Saxby, 1989; Badersten *et al*, 1984, Abbas *et al*, 1982), radiographic interpretation (Flack *et al*, 1996; Gröndahl, 1979), and notably restoration replacement (Poorterman *et al*, 1999; Pitts, 1997; Baders and Shugars, 1993; Kidd *et al*, 1993; Mjör and Toffinetti, 1992; Tveit and Espelid, 1992; Drake *et al*, 1990; Marynuik, 1990; Kay *et al*, 1988;

Braun and Marcus, 1985; Merrett and Elderton, 1984; Elderton and Nuttall, 1983; Bailit *et al*, 1979). Another area of particular difficulty (evident in both clinical and laboratory studies) is caries assessment (Ekstrand *et al*, 1997; Kidd *et al*, 1993a; Downer, 1989; Rytömaa *et al*, 1979; Rugg-Gunn and Holloway, 1974; Davies and Caldell, 1963; Berggren and Welanders, 1960; Slack *et al*, 1958; Knapp, 1868) and this is complicated by the presence of dental restorations (Foster, 1994; Kidd, 1989). Further research shows that the outcome of the investigation is made no more consistent or reliable when visual clinical examination is supplemented with radiographs (Ricketts *et al*, 1995, Ketley and Holt, 1993). This lack of agreement is a cause of concern to the dental community, the general public and those who fund dental treatment. This is not surprising as the cost of restoration replacement is huge (Mjör, 1993); US\$ 5000m (USA, 1984), NLG 600m (Netherlands, 1988), GB £104m (UK, 1991), (Chadwick *et al*, 2001; Jokstad *et al*, 2001). The variation in restoration replacement decision making (which is often based on subjective diagnoses) suggests that there may be a significant number of potentially sound and serviceable restorations being replaced unnecessarily. Clinical measurements must be reproducible for results to be valid and clinical trials and epidemiological data acquisition programmes have used calibration and training programmes to good effect to overcome problems of validity (Lazarchik *et al*, 1995; Horowitz *et al*, 1973; Davies and Caldell, 1963; Slack *et al*, 1958). It is self-evident that while clinical diagnostic tools are tested in a research environment they must be transferable, wherever possible, to the clinical setting to improve patient care. Ideally, restoration failure should be measured using simple, reliable and validated tools that can be used easily by practitioners.

The United States Public Health Service (USPHS) criteria were developed for use after the Californian Dental Association set up a Task Force on Quality Evaluation (Ryge and Snyder, 1973). These criteria evaluate restoration aesthetics, marginal adaptation and discoloration, anatomical form and recurrent caries using two independent and calibrated examiners. This tool has been shown to be workable in the field (Ryge *et al*, 1974; Eames *et al*, 1974; Osborne and Gale, 1974a and b; Osborne *et al*, 1973; Phillips *et al*, 1973).

As will be highlighted in the literature review, research into restoration replacement is full of inconsistencies and it is difficult to carry out realistic comparisons of what has been published as differences in study design, outcome measure or reporting method makes this difficult.

It is not surprising that there is a call for consistency in decision making in today's environment where spiralling costs and consumerism take on ever increasing and more significant roles. One starting point in helping to allay such concerns may be in taking steps to make sure that restorations are replaced only when it is felt absolutely necessary. There is a potential need to make sure that the consistency in decision making amongst practising dentists is acceptable and that ideally the consistency is not only comparable within a practitioner but between practitioners too. The development of the USPHS criteria over the years to evaluate work of dental auxiliaries appears to be a good starting point for such measures.

The purpose of this study was to determine the effect that restoration evaluation training had on the decision making and restoration replacement rates amongst a group of practising dental practitioners. As far as the author is aware this is not something that has been attempted before.

CHAPTER 2

REVIEW OF THE LITERATURE

2.1 Presenting the challenges

Two events play a major role in the practice of restorative dentistry: the decision to place a restoration and, subsequently, the decision to replace it (Mjör *et al*, 2005).

It is estimated that 70 to 80% of a dentist's workload is the placement, assessment and replacement of dental restorations (Mjör, 1981) and it is accepted that, despite technical excellence, **all** dental restorations **will** fail over a period of time (Jokstad *et al*, 2001). The failure of dental restorations is of major concern in dental practice and this is highlighted in the literature (Manhart *et al*, 2004). The failure of dental restorations can be a significant problem to patients, dentists and those who fund dental treatment. Ultimately, one of two outcome measures can be used to measure the longevity of a dental restoration; how long it survives until it fails *or* how long it lasts until it is replaced.

These outcome measures are distinctly different and are discussed later.

2.1.1 Defining failure

Failure is ¹ – “a thing that is unsuccessful or disappointing, not reaching the required standard in an examination.” The failure of a restoration, for whatever reason, may result in its replacement. Replacement is ¹ – “to substitute a thing for another, to take place of or supersede”. In dentistry it can be difficult to determine failure categorically (this is discussed later) and it is important to appreciate that the reasons for restoration replacement are not necessarily the same as the reasons for restoration failure. Clinicians need to appreciate that despite being inextricably linked restoration failure and replacement are not necessarily synonymous. Replacement may be subsequent to failure but it is not always clear

¹ Collins English Dictionary, Harper Collins Publishers, PO Box, Glasgow G4 0NB, 1993

cut e.g. technically sound restorations are replaced occasionally for aesthetic reasons rather than mechanical failure, equally a failed restoration resulting in the loss of a tooth may or may not be replaced with a dental prosthesis.

In search of better descriptive terminology, Lucarotti *et al* (2005a) coined the term “re-intervention” with respect to restoration replacement. From a clinical view and with respect to giving patients an idea of restoration life expectancy re-intervention rates appear to not only make sense but also, to a degree, simplify reporting. Re-intervention could include all instances when the history of the tooth or restoration is known and it would not be necessary to define failure or indeed the reason for re-intervention; it could be following a dentist’s decision or a patient’s request - it would make no difference to the re-intervention time as the restoration would have been replaced. It should be possible to calculate re-intervention rates prospectively or retrospectively from all types of clinical studies; randomised controlled trials, cohort studies, cross sectional studies from case notes or as in the case of Lucarotti *et al*’s work (2005a) from data base analysis. The calculation of such rates would allow comparisons at local and global levels and if significant differences appeared then further analysis through well-conducted research into the area might be justified.

2.1.2 The reasons for restoration replacement

Caries, tooth fracture, restoration fracture, deterioration in colour, loss of the previous restoration, poor anatomical shape, pain or sensitivity, a decision to change the original restorative material, or a desire to achieve the ideal have all been proposed in the literature as reasons for restoration replacement (Table 2.

1). Generally, there are three broad reasons for restoration replacement; material

failure, operator error and patient factors (Manhart *et al*, 2004). With reference to restoration replacement, at least one paper every decade or so has been published which corroborates previous researchers findings (Mjör and Gordon, 2002; Clarkson *et al*, 2000; Deligeorgi *et al* 2000; Burke *et al*, 1999; Mjör, 1997b; Gruythusen *et al*, 1996; Friedl *et al*, 1995; Friedl *et al*, 1994; Pink *et al*, 1994; Jokstad and Mjör, 1991b; Allander *et al*, 1990; Qvist *et al*, 1990a and b; Mjör, 1989; Mjör, 1986; Marynuik and Kaplan, 1986; Klausner and Charbenau, 1985; Marynuik, 1984; Hesselgren and Thylstrup, 1982; Mjör, 1981; Dahl and Eriksen, 1978; Lavelle, 1976; Richardson and Boyd, 1973; Moore and Stewart, 1967; Healey and Philips, 1949). The reasons for amalgam replacement have remained relatively consistent over the decades (Manhart *et al*, 2004; Mjör *et al*, 2000; Pink *et al*, 1994; Mjör and Medina, 1993; Mjör and Toffenetti, 1992; Klausner *et al*, 1987). However, the same cannot be said for composite resin as recurrent caries as a reason for replacement has decreased over time (Gaengler *et al*, 2004; Manhart *et al*, 2004; Pallensen and Qvist, 2003; Hickel and Manhart, 2001; Hickel *et al*, 2000). Inconsistency in restoration replacement decision making will always be present if the reasons for restoration replacement are multi-factorial and the decisions made in a subjective manner. By considering how failed restorations can be identified clearly through the objective assessment of their quality it should be possible to reliably identify the point at which a restoration should be replaced.

In addition to the points made above, it is also important to consider the effect that the replacement of a restoration has on a tooth's vitality and its survival. Generally, replacement restorations do not last longer than the original (Elderton, 1977b), contain defects (Elderton, 1977a) and, invariably, necessitate the further removal of tooth substance (Elderton, 1979). Therefore, if the downward spiral of

tooth morbidity (or tooth loss associated with restoration replacement) is to be avoided it would make sense that interventive procedures be carried out only when necessary.

2.1.3 The prevalence of restoration replacement

In the United Kingdom, over one-half of the dental restorations made in the National Health Service are replacements (Clarkson *et al*, 2000; Burke *et al*, 1999; Wilson *et al*, 1997a and b; Nuttall and Elderton, 1983). Similar figures have been reported in other countries: Australia (Tyas , 2005; Brennan and Spencer, 2003), Brazil (Braga *et al*, 2007), Canada (MacInnis *et al*, 1991; Boyd and Richardson, 1985; Richardson and Boyd 1973), Denmark (Qvist *et al*, 1990a and b, Qvist *et al*, 1986 a and b), Finland (Palotie and Vehkalahti 2002), Germany (Friedl *et al*, 1994), Greece (Deligeorgi *et al*, 2000), Iceland (Mjör *et al*, 2002), Italy (Mjör and Toffenetti 1992a and b), Jordan (Al Negrish 2001), Korea (Mjör and Um, 1993), the Netherlands (Bronkhorst, 1988), Sweden (Mjör 1997a) and the USA (York and Stephen, 1993; Marynuik, 1990; Klausner and Charbeneau, 1985; Douglass and Gammon, 1984; Moore and Stewart, 1967) (Table 2.2).

2.1.4 Estimating how long a restoration will last

As mentioned, two outcomes are used to quantify restoration failure in clinical research; failure or replacement. In order to make sense of published results the outcome measure used needs to be clearly stated. Replacement of a restoration is **not** always a specific measure of failure *per se*, it is most often a subjective decision made by a clinician, and this subjectivity is something that cannot be easily quantified or qualified. Studies that base their data collection on such

subjective information do not allow for an easy comparison of data and a significant part of the inability to answer clearly the question “How long will restorations last?” lies with problems relating to study design, data collection, data interpretation and data analysis that appear in the literature. Such problems are highlighted by comparing clinical trials of amalgam and composite with retrospective data from general dental practice. It being noted in clinical trials that composite restorations perform as well as (Mair, 1998) or even slightly better than amalgam restorations (Mjör and Jokstad, 1993; Smales *et al*, 1991a). These findings are not repeated in cross-sectional retrospective studies based in general dental practice, which suggests amalgam restorations survive longer than composite resin restorations (Mjör *et al*, 2000; Mjör, 1997a; Jokstad *et al*, 1994) - 6.6 to 14 years compared with 3.3 to 4.7 years. The discrepancies may be due to differences in working practice between the two environments, patient differences, or calibration and training differences; these “challenges” are dealt with in detail later. Additionally, there is the problem that many studies highlight the difficulty in getting dentists to agree (Poorterman *et al*, 1999; Pitts, 1997; Baders and Shugars, 1993; Mjör and Toffinetti, 1992a and b; Tveit and Espelid, 1992; Drake *et al*, 1990; Marynuik, 1990b; Braun and Marcus, 1985; Merrett and Elderton, 1984). This has particular relevance when investigating restoration replacement (Foster, 1994; Kidd, 1989) but it is present in other areas e.g. treatment planning (Kay *et al*, 1992; Grembowski *et al*, 1991; Elderton and Nuttall, 1983; Bailit and Clive, 1981; Bailit *et al*, 1979; Rytömaa *et al*, 1979), periodontal assessment (Walsh and Saxby, 1989; Badersten *et al*, 1984; Abbas *et al*, 1982), radiographic interpretation (Flack *et al*, 1996; Gröndahl, 1979) and caries assessment (Ekstrand *et al*, 1997; Kidd *et al*, 1993 a and b; Downer, 1989). It is difficult to say how long a particular

restoration will last, as there are many factors to take into consideration. However, it is probably safe to say that, in general, restorations should last longer today than they did in the past. The reasons are multi-factorial and include: improvements in a population's oral health (Nuttall *et al*, 2001; Hugoson *et al*, 2000; Kalsbeek *et al*, 2000; Schuller and Holst, 1998; Berset *et al*, 1996; Winn *et al*, 1996), increases in preventive dentistry measures, improved clinical techniques (Manhart *et al*, 2004; Hickel *et al*, 1998) and improvements in dental material technology (Hickel *et al*, 1998; Mjör, 1997).

2.1.5 The cost of restoration replacement

The cost of restoration replacement is huge (Mjör, 1993). Its cost has been estimated in some countries; US\$ 5000m (USA, 1984), NLG 600m (Netherlands, 1988), GB £104m (UK, 1991), (Chadwick *et al*, 2001; Jokstad *et al*, 2001). In today's climate where value for money and evidence based decision-making are watchwords, it would be beneficial to show that replacement dentistry is carried out appropriately and consistently and that dental practitioners make correct and wise decisions. If the decisions are correct then the cost of replacement dentistry may be viewed as justifiable. Equally important is the reliability of the decision-making process itself that should also be valid and repeatable, if only to satisfy interested third parties. Bearing these points in mind, even small reductions in restoration replacement decisions could make significant savings with respect to the financial outlay needed to sustain the spiral of restoration replacement.

2.1.6 Variation in decision-making

When examined closely it is noticeable that restorations are very often replaced for what can only be described as subjective rather than objective reasons. The diversity and variability of subjective assessment in dentistry is well-documented (Poorterman *et al*, 1999; Pitts, 1997; Boyd, 1989; Kennon, 1989; Hocott, 1984; Elderton and Nuttall, 1983; Milgrom *et al*, 1981; Natkin and Guild, 1967; Abramowitz, 1966; Gruebbel, 1950).

Bader and Shugars (1993) highlighted the problem in decision-making particularly well in their research based on the treatment recommendations from 51 dentists about 1,187 teeth in 43 patients. These researchers showed that the variation in decision making between dentists was complicated by the presence of a restoration in the tooth being examined. Inter-examiner agreement was lowest when previously restored teeth were evaluated and agreement correlations (measured by the kappa statistic) fell from 0.62 to 0.52² when teeth with and without restorations were compared. They also showed that a tooth housing a restoration was more likely to be scheduled for treatment. These findings, in view of the high cost in delivering “repeat” dental services (as noted above) is, therefore, of significance.

2.1.7 Trying to determine failure

As highlighted, rather than measuring failure, many studies measure when the restoration is replaced. This is particularly common in cross sectional surveys that unfortunately do not define what “replaced” means. In effect they do not have validated, objective or for that matter reliable end points in the decision making

² According to Landis and Koch (1977) these values relate to strength of agreement bands; 0.41-0.60 representing moderate agreement and 0.61 – 0.8 substantial agreement.

process; the decisions very often being made on subjective grounds with each dentist involved in the study working to their own inconsistent and undefined criteria.

Ideally, the measurement or determination of restoration failure should be simple, valid and reliable using measures that lack internal or external examiner variability. The use of objective outcome measures in general dental practice could allow a realistic evaluation of restoration longevity and overcome many of the problems noted in cross sectional surveys, case control and cohort studies and the randomised controlled trial; this is dealt with in Section 2.2.

Despite the problems alluded to, failure can be easy to determine in some cases e.g. if a restoration is

- completely lost,
- obviously fractured,
- clearly deficient

or

- unacceptable on aesthetic grounds.

Such “catastrophic” failures are easier outcomes to validate as long as there is agreement as to what represents the failure. Catastrophic failure often necessitates intervention but unless care is taken, it is possible that a failed restoration can be replaced in a study and the data neither collected nor reported. Unfortunately, as many clinical studies operate under strict budgetary limitations and take place over relatively short periods (1 or 3 years being the most common) catastrophic failure points may not be reached. However, clinicians by nature try to identify failing restorations before catastrophic failure occurs and often rely heavily on their own subjective opinion.

Attempts to define and quantify failure in dentistry have been made and Kreulen *et al* (1998) did this by separating failures into true and false failures (Table 2.3). Their work attempting to clarify between what can be attributed to as a cause of failure e.g. improper carving (Crawford, 1938) and that which is an effect or manifestation of the problem e.g. restoration loss, as this affects how findings may be reported i.e. failure due to technique or failure due to material inadequacies. Kreulen *et al*'s (1998) paper highlighted a number of challenges in defining failure and the difficulty in translating such findings to restoration replacement. For example, recurrent caries is difficult to diagnose with certainty (Fontanna, 1995; Kidd and O' Hara, 1990) and without clear guidance, dentists apply their own clinical thresholds to restoration replacement. Despite this, recurrent caries is reported as the commonest reason for restoration replacement in dental practice based research (Mjör and Gordan, 2002; Fontana and Gonzalez-Cabezas, 2000; Mjör, 1997b; Friedl *et al*, 1995; Friedl *et al*, 1994; Pink *et al*, 1994; Kidd *et al*, 1993; Mjör and Moorehead, 1993; Mjör and Um, 1993; Mjör and Toffenetti, 1992a and b; Qvist *et al*, 1990a and b; Allander *et al*, 1990; Kidd and O' Hara, 1990; Klausener and Charbeneau, 1985; Kelsey *et al*, 1981; Mjör, 1981; Skogedal and Helöe, 1979; Dahl and Eriksen, 1978; Moore and Stewart, 1967; Healey and Phillips, 1949). However, the incidence of recurrent caries in long-term clinical trials is only in the order of 1-2% over a fourteen to fifteen year period (Akerboom *et al*, 1993; Osborne *et al*, 1991). Such a finding gives credence to our need to establish what the effects of training in restoration evaluation have on the replacement rates of dental restorations. In addition, it should be established whether the variation in diagnosis between the real world and that in clinical trials is not just down to subject selection.

Evidence suggests that indirect methods of restoration evaluation (models or photographs) produce more critical evaluations than those achieved by clinical examinations alone (Johnson *et al*, 1992; Smales and Creaven, 1985; Goldberg *et al*, 1984; Smales, 1983; Mjör and Ryge, 1981). However, laboratory based evaluations often measure technical excellence and this does not necessarily equate with clinical success. Additionally, the use of laboratory-based tools has limited application in the clinical environment. Therefore, despite the trials, tribulations and pitfalls in taking a clinical approach to the assessment of restoration failure it is believed that this is the only way to accurately reflect the clinical performance of dental materials and dentists alike (Jokstad and Mjör, 1991a). It is also obvious that for a clinician to make a valued and respected judgement as to a restoration's likely future it needs to be based on an assessment system that uses reliable, familiar and readily available tools in a clinical environment (Jokstad *et al*, 2001).

2.1.8 Defining, determining and assessing quality of dental restorations

Defining and determining quality

Quality by definition (Oxford Popular English Dictionary, 1999) means "degree of excellence" but this differs from industrial standards that refer to quality as "a measure of a product against an objective standard." There is little agreement in the literature as to exactly what quality in dentistry means. This challenge in defining quality in dentistry is problematic (as has been noted with trying to define failure) and that, as stated earlier (Jokstad *et al*, 2001), there is no direct relationship between restoration longevity and technical excellence e.g. it is possible for a tooth to be restored with an apparently perfect restoration yet

present with symptoms of acute pulpitis. With reference to restorations, quality could pertain to quality of appearance, quality of performance, cost or indeed any of these in combination. In dentistry, as a whole, measurements of quality are generally used for specific purposes such as peer review (Schonfield, 1969), the assessment of work carried out by a third party (Friedman, 1972) and in order to determine the nature and standard of dental care provided in a population (Schonfeld, 1967).

It is possible to assess the quality of dental restorations directly and indirectly and a number of tools have been developed to assess specific features of dental restorations e.g. the margin (Grossman and Metejka, 1997; Roulet, 1988; van Amerongen and Eggink, 1986; Mahler and Marantz, 1979; Osborne *et al*, 1976), surface wear (Bryant and Hodge, 1994; Leinfelder *et al*, 1986) and surface roughness (Smales and Creaven, 1979). How these determinations relate to clinical performance and failure needs evaluation e.g. it is possible to measure with great accuracy how a composite resin wears over time but at what point is the wear great enough to affect the patient, or influence the occlusion?

Assessing quality

A number of strategies to determine the quality of dental procedures have been published (Abramowitz and Mecklenberg, 1972; Hammons *et al*, 1971; Lotzkar *et al*, 1971; Soricelli, 1970; Abramowitz, 1966). These methodologies differ in their approaches and were developed to achieve different aims. Lotzkar *et al* (1971) and Abramowitz (1966) developed their strategy to assess the quality of restorations placed by dental auxiliaries. Soricelli (1970) assessed the work of dentists and expanded function dental assistants in Philadelphia, USA.

Abramowitz and Mecklenberg (1972) used their tool to assess the work of dentists. Although used locally none of these assessment methods has been adopted on a global scale.

However, there is an international and widely recognised method for the assessment of restorations, namely the United States Public Health Services criteria (USPHS) (Ryge and Snyder, 1973). This tool is workable in the field (Ryge *et al*, 1974; Eames *et al*, 1974; Osborne and Gale, 1974a and b; Osborne *et al*, 1973; Phillips *et al*, 1973). The original Ryge criteria were developed for use after the Californian Dental Association (CDA) set up a Task Force on Quality Evaluation in 1973. Although this thesis is concerned with the specific area of the evaluation of dental restorations, the Task Group looked at a number of areas in dentistry. These included history and clinical examination, radiographs, diagnosis, treatment planning, management of pain, anxiety and emergencies, preventive dentistry, endodontics, oral and maxillofacial surgery, operative dentistry, crowns and fixed partial prosthodontics, removable partial prosthodontics, complete denture prosthodontics and orthodontics. The whole CDA Quality Evaluation system was published as a manual entitled "CDA Quality Evaluation of Dental Care: Guidelines for the Assessment of Clinical Quality and Professional Performance in 1977. The manual has been updated and can be located at the Californian Dental Association uniform resource locator (url)

http://www.cda.org/member_benefits_&_resources/peer_review/guidelines_for_the_assessment_of_clinical_quality_&_professional_performance

2.1.9 The USPHS criteria

Reasons for using USPHS

According to Jokstad *et al* (2001), only tools commonly available in dental practice should be used to determine the clinical acceptability restorations. The USPHS criteria lends itself to the routine assessment of dental restorations as it uses everyday standard dental equipment that all dental practitioners are familiar with i.e. a dental probe and mirror.

Despite other tools being available to assist the dental practitioner in clinical diagnosis e.g. operating microscopes (Haak *et al*, 2002), intra-oral camera (Forgie *et al*, 2003), the DIAGNODent laser device³, Digital Imaging Fiber-Optic Transillumination (DIFOTI)⁴ and quantitative light-induced fluorescence, these were not used in this research as they are not readily accessible to the “average” general dental practitioner. There is also a concern as to whether or not these new devices really do offer any significant advantage over the traditional examination (Pretty and Maupomé, 2004). As an example the electronic caries monitor exhibiting a sensitivity of 65% and specificity of 73% compared with a sensitivity of 60% and specificity of 73% achieved with a visual examination (Ashley *et al*, 1998) and also comparing this with the results from quantitative light-induced fluorescence (a sensitivity of 68% and specificity of 70%) (Pretty *et al*, 2003).

While the use of radiography in the assessment of dental caries has been shown to have some benefit over a basic examination in some population groups (Hintze and Wenzel, 2002; Hintze and Wenzel, 1999; Mileman and van der Weele, 1990) this is not always the case (de Vries *et al* (1990). It appears that, overall,

³ KaVO, Germany

⁴ www.difoti.com

the value of radiographic exposure and its diagnostic value is related to the incidence of caries within the population group being examined (Pretty and Maupome, 2004e). It also worth noting that research suggests that radiographic examination yields high certainty with respect to absence of caries (specificity = 0.97 or 97%) and only moderate certainty with respect to presence of caries (sensitivity = 0.54 or 54%) (Pretty and Maupomé, 2004c). Additionally, there is also the argument that radiographic exposure needs to be justified and not used to make clinical diagnoses without a thorough examination of the patient beforehand and an actual need for a radiograph determined (Pretty and Maupomé, 2004a). This point has also been made with respect to caries status around restorations (Espelid and Tveit, 1991). These researchers noting that with Class II amalgam restorations a false-positive return of 12% dropped to 3% when radiographic evaluation was supplemented with a clinical examination (the true-positive return increased from 47% to 64%).

It is felt that the clinician needs to maintain the responsibility for restoration replacement. The decision process itself is multi-factorial (this includes the influence of any patient involvement) and although there may be a major single factor that could necessitate restoration replacement it is equally possible that replacement could be considered after taking into consideration multiple minor defects. Pretty and Maupomé (2004a-e) suggest that clinicians should not base their decisions on a single diagnostic system but consider all relevant information then formulate a decision as to the best line of treatment to support this view reflect this philosophy.

There is, yet, no universally acceptable method adopted by the dental community for the assessment of dental restorations. However, the USPHS

system appears to be a popular tool in restoration assessment research and the Federation Dentaire International (1980) has adopted it for assessing the status of restorations (Mjör, 1986). Presumably, because the USPHS criteria lends itself to field work which is not necessarily the case when less well-defined evaluation criteria are used (Cardoso *et al*, 1989) and that it has also been shown to be beneficial over more detailed systems (Mjör and Haugen, 1976) despite some of its deficiencies which are highlighted.

Reasons against using USPHS

The USPHS criteria have been used for many years in clinical field trials (de Arujo *et al*, 1998; Thordrup *et al*, 1998; Abdalla *et al*, 1997; Hse and Wei, 1997; Abdalla and Aldahainy, 1996; Holan *et al*, 1996; Matis *et al*, 1996; Navarro *et al*, 1996; Qualtrough and Wilson, 1996; Wilson *et al*, 1996; Cipriano and Santos, 1995; Powell *et al*, 1995; Rasmusson and Lundin, 1995; Sjogren *et al*, 1995; Tidehag and Gunne, 1995; Wendt and Leinfelder, 1994; Granath *et al*, 1992; Wendt and Leinfelder, 1992; Matis *et al*, 1991; Wilder *et al*, 1991; Wilson *et al*, 1991; Stangel and Barolet, 1990; Wendt and Leinfelder, 1990; Brunson *et al*, 1989; Capel *et al*, 1989; Lundin and Koch, 1989; Cavel *et al*, 1988; Wilson *et al*, 1988; Oldenburg *et al*, 1987; Sturdevant *et al*, 1986; Davis and Mayhew, 1986; Beere *et al*, 1984; Straffon *et al*, 1984; Paquette *et al*, 1983; Hamilton *et al*, 1983; Leinfelder *et al*, 1980; Tonn *et al*, 1980). However, the fact exists that the assessment criteria have never been validated with respect to a definitive need for restoration replacement. Although USPHS essentially “bands” restorations into various categories there is and can be no absolute guarantee that if left alone that a restoration could not function well for an immeasurable time. Arguably, there

would be an ethical dilemma to be overcome if clinically deficient restorations were left unmodified, for whatever reason, once they had been diagnosed as requiring replacement. For this reason alone, there is much to be said for the commitment to introduce a reliable and deliverable restorations assessment programme whether this is through the modification of the USPHS criteria or the introduction of a completely new system (Hickel *et al*, 2007). The implications of practice-based research and the networks required to carry out such an initiative cannot be underestimated (Green, 2001). Especially when there are key issues that need to be overcome (Bader and Shugars, 1992; Tveit and Espelid, 1992; Espelid and Tveit, 1991; Tveit and Espelid, 1986);

- treatment decision variability by clinicians
- lack of standardisation and calibration used for decision making
- what exactly constitutes failure?

As a research tool, the USPHS system lacks measurement sensitivity in short term clinical investigations as the assessment criteria often fail to identify the failing restoration during its early years of service (Hickel *et al*, 2007). It also suffers from being discrete and relative rather than an absolute grading system that does not allow the researcher to apply routine parametric statistical analyses to its results. Such a limitation, means that if we are to measure a specific parameter e.g. restoration deterioration in plastic filling material over a period of time, then other evaluation tools are required. These exist for some parameters e.g. when measuring wear and it would appear from the literature that indirect measurements could provide us with the gold standard in this area. Leinfelder (1985) and Goldberg *et al* (1984) provide typical techniques for the measurement of restoration wear and these have been shown to have precision and sensitivity with

good inter- and intra- examiner agreement when they are used and applied correctly (Taylor *et al*, 1984). However, while these tools are used with relative ease in the laboratory, their application and relevance to the clinical situation has to be considered; other techniques are available in the assessment of materials and they too have their applications in research e.g. the use of clinical photographs or scanning electron micrographs.

Arguably, over the past 35 years, the clinical performance of dental materials has improved markedly and as mentioned above the USPHS criteria lack sensitivity for short-term evaluations of clinical materials. Therefore, there, may be a call for the introduction of a different and more robust tool for the short term (up to 2 years) evaluation of dental materials in controlled clinical trials (Hickel *et al*, 2007). It being noted that “the majority of restorations in many clinical studies continue too receive an Alfa score at 6, 12 and 18 months”. While the shortcomings of some clinical studies i.e. study design, recruitment of subjects, numbers of restorations per subject and operational procedures have been highlighted in the literature review it is worth discussing and detailing a number of shortcomings with the individual evaluation criteria within the USPHS.

Ultimately, the problem with any analysis is whether statistical significance equates with clinical relevance. A similar problem exists when we consider the clinical relevance of research that uses questionnaires to measure the reasons for restoration replacement (Boyd and Richardson, 1985; Klausner and Charbeneau, 1985; Mjör, 1981). There is often little in the way of validating the replacement decisions which are reported and to all intents and purposes they should be considered extremely subjective. While such studies can provide us with details on actual clinical restoration replacement their overall validity and applicability to other

populations is questionable. They may rather reflect the prescribing patterns of the practitioner.

Using USPHS

The USPHS criteria requires, in its strictest sense, the use of two independent examiners with the system being based on evaluations of restoration aesthetics, marginal adaptation and discoloration, anatomical form and recurrent caries. Each parameter has a range of scores to categorise a restoration and these range from Alfa to Delta (depending on the evaluation parameter); Appendix 2.1 summarises how the scores for particular parameters are made.

As mentioned, the calibration of the examiners is a prerequisite to the successful use of the USPHS criteria. Unfortunately, calibration programmes are not always undertaken (Mjör and Gordon, 2002). When calibration has not been undertaken, the validity of such studies' findings has to be questioned. The examiners will have used their own judgement that will be different to others or indeed with themselves at different times of the study. The value of calibration and standardisation in assessment is particularly relevant in the evaluation and determination of restoration failure as variation between examiners leads to inconsistent recommendations on replacement; the public does not look upon such variation favourably.

The potential challenges in applying the USPHS criteria to restoration replacement is highlighted by considering the relative validity of each of the individual component parts of the USPHS criteria.

Assessing caries

Without doubt, the presence of caries around a restoration could be considered as reasonable justification for restoration replacement, even though it is possible that no intervention, or active remineralisation / caries control measures, could result in a lesion remaining static and the tooth continue to function admirably for many years. The presence or absence of caries in a tooth is something that produces inconsistent examiner agreement in dentistry and it is well recognised that practitioner's decision making in caries diagnosis is, at best, variable (Rytömaa *et al*, 1979; Todd, 1975; Davies and Caldwell, 1963; Berggren and Wellander, 1960; Slack *et al*, 1958). Caries status can be diagnosed with near 100% accurately from the histo-pathological analysis of extracted and sectioned tooth. Nevertheless this is not something that can be readily utilised in the clinic - the histo-pathological assay of teeth to determine caries status is not and will never be accepted as a clinically acceptable technique. To this end, practitioners have to decide themselves (they do not necessarily have the luxury of calling upon a colleague to validate their findings) whether or not they feel caries is present in a tooth and, if it is found, does its presence have significant bearing with respect to a need for intervention or treatment through restoration replacement. As mentioned, the detection of caries associated with the margin of a restoration is not easy to determine (Goldberg, 1990) and marginal discolouration and restoration defects can be inadvertently labelled as sites for recurrent caries (Mjör, 2004) and commit restorations to the re-restoration spiral. The lack of objective criteria and suitable procedures for diagnosing recurrent caries is lacking in dentistry which commits significant numbers of teeth to re-restoration when there is no real need; Söderholm *et al* (1989) determining with histology that only 37% of sites

diagnosed with recurrent caries are in fact carious. These researchers also showed in their study that around 10% of carious lesions were missed on evaluation. While radiographs may have value in the detection of cervical lesions (Hewlett *et al*, 1993; Espelid *et al*, 1991), they only have diagnostic value if the x-ray beam passes directly through the defect and not masked by the presence of a radio-opaque restoration (Tveit *et al*, 1991). The difficulties associated with the diagnosis of recurrent caries and the lack of examiner calibration in clinical evaluations reduces the diagnostic validity of many research projects as caries is likely to have been misdiagnosed. This potential for misdiagnosis is particularly relevant to clinical trials or cross-sectional studies which set out to evaluate the reasons for restoration replacement (Manhart *et al*, 2004; Hickel and Manhart, 2001; Hickel *et al*, 2000). An international committee has been formed with a purpose to improve caries diagnosis (ICDAS, 2005). Their recommendations will have particular relevance to the diagnosis and recording of recurrent caries in the future; at present long term longitudinal research (> 10 years) reports low rates of occurrence i.e. 4 - 8% (Gaengler *et al*, 2004; Manhart *et al*, 2004; Pallesen and Qvist, 2003; Hickel *et al*, 2001; Manhart and Hickel, 2001; Hickel *et al*, 2000).

Assessing marginal discolouration

Although marginal discolouration is used as a potential indicator of caries presence in a tooth, it is a poor measure for determining such pathology (Rudolphy *et al*, 1995). It is, possible, for example, that marginal discolouration may be consequent or even secondary to the use of certain materials e.g. dental amalgam. It is important that evaluators consider this and only measure discolouration beyond that normally expected with the restorative material used.

The problem of marginal discolouration is not exclusive to amalgam and there is diagnostic difficulty in associating stained and discoloured margins around tooth coloured restorations with recurrent caries (Kidd, 2001; Kidd and Beighton, 1996; Kidd *et al*, 1995; Tyas, 1991; Kidd, 1989).

Assessing marginal integrity

The validity of marginal gap size is a poor indicator of restoration quality, restoration longevity or when establishing recurrence of caries with plastic restorations (Mjör and Toffenetti, 2000; Kidd *et al*, 1995; Pimenta *et al*, 1995; Hewlett *et al*, 1993; Kidd and O'Hara, 1990; Goldberg, 1990; Söderholm *et al*, 1989; Merrett and Elderton, 1984). Although some research has shown that there is an increased likelihood of recurrent caries when the marginal gap size exceeds 35µm (Goldberg, 1990; Tveit, 1990; Goldberg *et al*, 1981; Jørgensen and Wakumoto, 1968) it is still an area that attracts controversy. More recently, it has been suggested that recurrent caries is not associated with crevice size around a restoration until we reach what is termed as "macro-leakage" and that this equates to between 250 and 400µm (Kidd *et al*, 1995). Additionally, "ditching" which is often found on the occlusal cavo-surface margins is not associated with recurrent caries (Mjör and Qvist, 1997; Mjör, 1985). The laboratory evaluation of marginal integrity relies on the evaluation of dye penetration (Going, 1972) or the two dimensional examination of scanning electron micrographs of restoration margins (Roulet *et al*, 1989). Both techniques have been criticised for having poor correlations with clinical performance (Heintze, 2007; Gaengler *et al*, 2004; ISO, 2003; Opdam *et al*, 1998; Noak *et al*, 1995). Although, it has been reported that restorations with marginal deterioration and cavo-marginal discolouration are more

likely to fail at five years than those without defects at three years (Hayashi and Wilson, 2003). Overall, the evidence on the influence that marginal integrity has on restoration failure rate is inconclusive. Again, there has been a call for future clinical studies to develop a scientific basis to assess this criteria in order to support or refute such a relationship in restoration survival (Hickel *et al*, 2007).

Assessing anatomical form

The validity of relating anatomical form to restoration longevity can be questioned. While it is possible that incorrect or insufficient restoration anatomy can lead to occlusal changes in the dentition the influence that an anatomical defect on single tooth can have on its likely failure in the long term cannot be fully validated.

Assessing aesthetics

Unlike the areas mentioned above, the aesthetic or perceived quality of a restoration's aesthetics could be validly used as a measure, an indicator or a "justifier" for restoration replacement. Validation being achieved if the patient and the dentist agree that a restoration is not fit for purpose and that it should be replaced on aesthetic grounds whether or not it is deficient in other areas.

As a reference standard for restoration integrity USPHS remains held in high regard. It has been said that it is the "only acceptable method for the assessment of restorations" (Mjör and Gordon, 2002) and while its measurement validity can be questioned in the absence of any superior methodologies it remains, perhaps, "the best tool for the job". It has to be recognised that there are on occasions no *bone-fide* biologic tests or assays that can be used to fully determine disease or disease conditions e.g. temporomandibular-joint pain

dysfunction syndrome, or indeed useful reference standards for the measurement of disease activity e.g. active disease in periodontal pockets.

2.1.10 Calibration in clinical research

As covered earlier (Sections 2.1.4, 2.1.6) many clinical decisions rely on subjective diagnoses and consequently a significant difference in decision-making occurs between clinicians in many areas of dentistry. Clinical and laboratory studies show that the assessment of caries by dentists is unreliable (Ekstrand *et al*, 1997; Kidd *et al*, 1993; Downer, 1989; Rytomaa *et al*, 1979; Rugg-Gunn *et al*, 1974; Davies and Caldell, 1963; Berggren and Welander, 1960; Slack *et al*, 1958; Knapp, 1868) and that caries diagnosis is further complicated by the presence of dental restorations (Foster, 1994; Kidd, 1989) and that the examination findings are not necessarily enhanced if it is supplemented with radiographs (Ricketts *et al*, 1995, Ketley and Holt, 1993).

One area that is particularly susceptible to inter-examiner variability is the decision about when to replace a restoration (Poorterman *et al*, 1999; Pitts, 1997; Baders and Shugars, 1993; Kidd *et al*, 1993; Mjör and Toffinetti, 1992; Tveit and Espelid, 1992; Drake *et al*, 1990; Marynuik, 1990; Kay and Watts, 1988; Braun and Marcus, 1985; Merrett and Elderton, 1984; Elderton and Nuttall, 1983; Webster and Mink, 1981; Bailit *et al*, 1979). Such lack of agreement between dentists can be seen as a cause for concern, it being a constant source of aggravation within the dental community and the public over the years.

Arguably such discrepancies in clinical decision making may be made worse by a lack of formal training or calibration within the undergraduate dental curriculum –

“... calibration in clinical judgement of restorations is seldom included in the dental curriculum” (Mjör et al, 2004).

By failing to introduce such measures in the formative years of the dental curriculum it is possible that importance of calibration as a whole is not being reinforced forcefully enough.

In order to reach acceptable levels of inter-examiner agreement examiners taking part in clinical trials and epidemiological data acquisition programmes undertake pre-examination calibration and training (WHO, 1977; Shaw and Murray, 1975; Horowitz *et al*, 1973; Davies and Caldell, 1963; Slack *et al*, 1958). Such calibration programmes have been shown to be workable and effective at improving examiner consistency and reliability (Gröndahl, 1979; Davies and Caldell, 1963; Slack *et al*, 1958). The use of fewer than four examiners reduces examiner disagreement in clinical evaluations (Slack *et al*, 1958). What we do not know is how much the variation decreases if they are all trained to use an index in the same way.

Despite the success of some calibration programmes there are instances where attempts to calibrate examiners has proved unsuccessful or only partially successful (Robertello and Pink, 1997; Scruggs *et al*, 1989; Poulsen *et al*, 1980; Hinkelman and Long, 1973; Fuller, 1972). By contrast, there are instances where success in clinical evaluation and re-evaluation has occurred without any examiner training (Goepford and Kerber, 1980; Mjör and Haugen, 1976; Abou-Rass, 1973). The variation in treatment provision between un-calibrated and calibrated examiners is significant (Rytömaa *et al*, 1979; Shaw and Murray, 1975; Horowitz *et al*, 1973; Davies and Caldell, 1963; Slack *et al*, 1958) and this may have a direct

bearing on treatment provided (Rytomaa *et al*, 1979) with both over-prescription and under-prescription of treatment being likely (Bader, 1992).

The dental community generally agrees that in order for research and research findings to be of the highest standard calibration programmes are not only desirable but in most instances essential. If clinical measurements are not reproducible and valid then the inferences drawn from them can only be considered as suspect. It is also self-evident that diagnostic tools need to not only be applicable to the research environment but also transferable to the clinical setting, be that within the primary or secondary care sectors.

The most effective way of delivering calibration programmes has not been established and the best approach to restoration assessment undetermined. It has been suggested that calibration courses should occur at research meetings and include re-calibration procedures (Hickel *et al*, 2007).

2.1.11 Summary

It is difficult to measure, assess, qualify and quantify the reasons for restoration failure and replacement (Elderton, 1976a and b) and the literature identifies a number of areas for concern:

- the number of restoration replacements carried out throughout the world,
- the subjectivity of the decision making processes with respect to restoration evaluation,
- the difficulties faced in trying to get dental practitioners to agree with each other.

The above highlights the need for research into restoration replacement decision making and how the application of structured and objective assessment processes

could improve the diagnostic process. It is important, to ascertain the actual life expectancy of restorations in dental practice. If the life expectancy of a restoration can be correlated with failure and replacement then it may be more possible to advise on and to answer the question “How long does a restoration last?”

2.2 How long does a restoration last? An evaluation of the literature by examining study design

Clinical studies can be classified into one of three major groups;

- cross-sectional studies
- cohort studies
- and*
- clinical trials.

The following section deals with the challenges faced in trying to determine how long a dental restoration lasts when different failure or survival criteria are used and the value that can be placed on the findings when we examine the three major study designs listed.

When analysed, many studies looking at restoration replacement are poorly designed, show variation in clinical procedure, use different materials or utilise unique evaluation criteria to the material being evaluated (Chadwick *et al*, 1999; Downer *et al*, 1999; Jacobsen, 1988). These variations in study design can create insurmountable problems when it comes to data comparison. The variation inherent in clinical research is reflected in the significant number of papers being classified as “unsuitable” for inclusion into two systematic reviews that attempted to evaluate the longevity of dental restorations; over 250 exclusions by Chadwick *et al* (2001) and 66 exclusions by Downer *et al* (1999).

It is difficult to compare longitudinal clinical trials - that are invariably carried out under strict research protocols on highly selected patients and in ideal clinical situations - with cross-sectional studies based in general dental practice (Mjör, 1997b; Roulet, 1997; Chadwick *et al*, 2001). There are many effect modifiers that need to be taken into consideration; clinical procedure, population characteristics, control groups, study environment, type of cavity, failure criteria, payment system and research sponsorship (these are discussed later).

Challenges like those mentioned have been recognised for many years;

“That some restorations fail is an undoubted fact, but the magnitude of the problem and the reasons for the failures are subjects for speculations since there have been no appropriately designed long-term studies to investigate these matters fully. Such studies would not only be difficult to undertake, but also they would inevitably be subject to considerable bias on account of patient selection and operator variation, both at the time of insertion of the restorations and later in their assessment. Nevertheless, there are published studies which yield useful information on the prevalence of failure of restorations...”

Elderton (1976a)

and

“co-ordinated research...in primary dental care...requires the establishment of multi-centre, multi-operator studies with stratification of tooth type, cavity type and other effect modifiers, for assessment periods of greater than ten years is needed if true answers on many aspects pertinent to restoration longevity are to be obtained e.g. a clinician’s skill, tooth type, type of cavity and material used”.

(Chadwick *et al*, 2001)

At the same time, the value of short-term clinical trials cannot be underestimated in helping to determine the reasons for restoration failure as they are often the only way in which performance differences between dental materials can be accurately tested and evaluated (Duke, 1992). Who would want to continue using a dental material that showed significant and early catastrophic failure? For example, in 1996 Navarro *et al* showed that 100% of the Gallium alloy restorations made in their study were replaced by 8 months; alarmingly, 3 of the 26 teeth restored suffered significant tooth fracture.

Adding to the difficulties is the fact that clinical practice is deeply influenced by the practitioner's ability to measure or predict a restoration's longevity (Mjör, 1989; Marynuik and Kaplan, 1986; Allan, 1977; Elderton, 1976a and b). Elderton (1983) illustrates this point particularly well. Out of 232 restored tooth surfaces 155 were scored as failures and scheduled for replacement by at least one of the fifteen dentists involved in the study. The variation between the dentists was staggering; one dentist planned to restore 20 surfaces while another proposed to restore 153.

Additionally, there is the problem that a dental restoration's survival is a reflected in material type e.g. amalgam versus composite resin, location e.g. smooth surface restorations versus fissure restorations, oral environment e.g. xerostomic with non-xerostomic patients and that restoration survival rates reflect material use and abuse in handling by dentists (Sano *et al*, 1998; Lussi *et al*, 1995; Billington *et al*, 1990; Mjör *et al*, 1990). Unfortunately, very few of these factors have been tested in clinical trials so it is extremely likely that a significant number of previously conducted clinical trials contain unknown confounders.

The problems identified are not new:

“no satisfactory, large-scale longitudinal data were available for the longevity of restorations, and that it was impossible to draw conclusions of general validity from published studies”.

(Marynuik, 1984)

Some twenty years later, little has changed. However, it is possible to group the literature and illustrate why Marynuik's (1984) statement still holds true.

2.2.1 Cross sectional studies and the challenges they pose

Cross sectional studies have attempted to quantify the longevity of dental restorations. This is particularly noticeable for amalgam and some of these studies are detailed in Table 2.4.

Cross sectional studies are, in essence, surveys carried out at a specific point or period in time. They have a number of strengths and weaknesses. Their greatest strength relates to the relative simplicity with which they can be carried out and that they can, when carried out correctly, be used to report the average view and practice of general dental practitioners (Mjör, 2004) and stimulate areas for future research. The strength of the report's findings is often related to the response rate of the survey.

A typical cross sectional survey may ask a dentist to report the reasons why they replaced ten consecutive restorations and generally, no guidance is given. This results in many cross sectional surveys presenting the subjective opinions of dentists applying pseudo-reasoning to the acceptability (or not) of a restoration they have replaced. Consequently, they are not a good tool for determining the life expectancy of a dental restoration. Additionally, cross sectional surveys do not report on what happens to successful restorations and this affects considerably

the objectivity of cross sectional surveys to determine the longevity of restorations in general (this is discussed later).

Challenges relating to sampling and response rates

It is not uncommon for cross sectional studies to fail to report their overall sample percentage (Mjör 1997b, Drake *et al* 1990, Klausner *et al* 1987, Qvist *et al* 1986a and b, Klausner *et al* 1985, Mjör 1981). The response rate to cross sectional surveys is variable; Al Negrish (2001) 88%; Mjör (1997) 71%; Klausner *et al*, (1985) 61%; Richardson and Boyd (1973) 56%, Pink *et al* (1994) 24%; Klausner *et al* (1987) 21%. The overall value of the results must be interpreted with caution until both the response rate and the size of the sample is verified.

A survey carried out in Canada by Richardson and Boyd (1973) represents a typical cross sectional study. This study utilised subjective reporting from a group of dentists (50 of 93) who recorded the number of and reasons why they replaced amalgam restorations over a period of five consecutive working days; twelve reasons for replacement were reported. The validity of this study must be questioned as to its generalisability as the participants represented less than five percent of the total number of dentists in Canada. Despite this, the authors calculated the number of restorations likely to be replaced by the average dentist in Canada and related it to a cost of CAN\$ 9,108,000 per year. Despite, as highlighted, flaws in sampling it is not unusual for data from cross sectional studies to be used to estimate the cost of restoration replacement to a “health service” or “dental care provider” (Jokstad *et al*, 2001; Smales *et al*, 1990). It is notable that the discussion sections in many cross sectional studies mention that the collected data might be unreliable yet still be used to highlight common findings from

general practice and that generally the authors felt their sample was representative of the particular country the survey was being carried out in.

Challenges relating to subjective data analysis

The majority of cross sectional studies do not use calibrated examiners and this is a problem. Evaluations carried out without training do not record failure but actually report subjective replacement decisions amongst dentists; in effect each dentist is working to their own set of criteria which may vary considerably.

Although the results are often portrayed as being typical of a practice setting they are in fact slightly divorced from such as the decisions made can be influenced by any number of things e.g. socio-economics, patient demography, dentist to patient ratio and variations in personal evaluation. Such variations between and within such studies raise doubt in the value of the results obtained (Jokstad *et al*, 2001).

This subjectivity gives an inherent unreliability to most cross sectional data.

However, variation may be reduced by ensuring that many operators complete the data collection process.

Some researchers attempt to reduce subjectivity in cross sectional studies looking at restoration failure by providing the participating dentists with explanation or categorisation of features that represent failure (Mjör *et al*, 2002; Al Negrish, 2001; Mjör, 1997; Mjör *et al*, 2000b; Drake *et al*, 1990, Qvist *et al*, 1986a and b). This approach has not been consistent in description or use. It is noticeable that even when given descriptions of the problems that there was still scope for personal interpretation by individual dentists if the descriptors used were not accompanied by training.

Despite their inherently subjective nature, cross sectional studies report many reasons as to why restorations are replaced by particular dentists or group of dentists. However, the findings between different countries are surprisingly similar despite different populations being examined and different payment systems operating in different health care systems e.g. caries being the most commonly reported reason for restoration replacement (Mjör *et al*, 2001; Mjör *et al*, 2000a and b; Burke *et al*, 1999; Mjör, 1997; Klausner *et al*, 1987). This finding is noted despite the uncertainty with which caries is diagnosed (Mjör and Toffenetti, 2000). This is something that is diametrically opposite to the findings in controlled clinical trials that fail to corroborate recurrent caries as the major reason for restoration replacement (Mair, 1998; Letzel *et al*, 1989).

Challenges relating to materials and /or clinical procedure

Lack of detailed data⁵ is a common finding in cross sectional studies. It is also notable that many cross sectional studies often try to investigate factors that are not always recorded accurately in clinical records. Cross sectional studies usually report generic material type and it is not often possible to determine differences between the materials reported on e.g. different dental alloy compositions.

Consequently, it is not possible to delineate or separate out the advantages and disadvantages of specific dental materials. Additionally, such research does not regularly identify differences in clinical technique e.g. rubber dam, local analgesia, etc. Despite this, cross sectional studies are often used to represent the failings of a restorative material and fail to account for operator factors (Osborne and Gale, 1974a).

⁵ It is accepted that this may be due to publishing constraints as well as being “missing” data

Challenges relating to the patient's past dental experience

Cross sectional studies pay scant reference to the patient's caries susceptibility or past dental experience. If caries is the most common reason for replacing a restoration (Friedl *et al*, 1994; Allander *et al*, 1990; Qvist *et al*, 1990a and b; Dahl and Eriksen, 1978; Mjör, 1981; Qvist *et al*, 1986a and b; Richardson and Boyd, 1973) (and this is correct) and dietary intervention is neglected by the patient and dentist then it is not surprising that recurrent caries may be among the commonest reason for subsequent restoration replacement.

Challenges relating to the time of sampling

Cross sectional studies report findings in relation to specific populations and at specific points in time. As a result, findings should not be extrapolated to other groups but limited to restorations of the same category and material. Some cross sectional studies indicate that tooth coloured restorations are replaced more often than amalgam restorations (Qvist *et al*, 1985; Mjör 1981). This is an example of an unfair comparison, as amalgam cannot be considered an aesthetic filling material. It would not make sense to replace one amalgam for another on the grounds of appearance whereas this may well be a reason for replacement with composite resins. However, it is noted that one study did not support the finding that tooth coloured restorations are replaced more often than amalgam (MacInnis *et al*, 1991).

Challenges relating to data manipulation

A study by Mjör *et al* (2000a) illustrates other common faults found in some cross sectional studies. Mjör and his co-workers reported on a cross-sectional study of

6,761 restorations in permanent teeth and made some interesting comparisons and conclusions. They suggested that male clinicians were less likely than female clinicians to be involved in the early replacement of restorations, that salaried dentists replaced restorations less than private practitioners and that replacements are more likely carried out by clinicians with the least clinical experience. These results arise, as is common with a lot of cross sectional studies, from internal comparisons between subgroups in the survey and although some interesting findings are noted, like those seen above, there is really no justification for the data to be used this way to make such comparisons. A more appropriate study design to determine such a difference would be an adequately designed cohort study.

A look at the original data presented by Mjör *et al* (2000b) immediately highlights some other problems. The original paper (Mjör *et al*, 2000b) reported on 22,391 restorations in permanent teeth. The subsequent paper based its analyses on 6,761 of the 11,800 replacement restorations carried out in permanent teeth; namely the restorations whose age at replacement could be determined and excluding nearly 43% of the restorations i.e. the ones that could not be aged.

The manipulation of data to find or search for findings of interest is not considered good practice and the value of applying this data to other patient groups limited (Crombie, 2005) yet it frequently occurs. It is also interesting to note that on occasion, although reported by authors, there is often no rational explanation of their findings e.g. Mjör *et al* (2000b) report no reason for the differences in replacement restoration rates for gender or salary as when analysed their reasons for replacement were the same for both groups (Mjör *et al*, 2000b; Mjör, 1999).

The value of cross sectional studies in evaluating restoration longevity

With respect to restoration replacement and longevity cross sectional studies only have value in determining what has been lost, what has failed (if failure criteria are defined) or what is being replaced at a specific point in time. They take no account of what happens to or is likely to happen to restorations that remain intact in a patient's mouth. This results in the accumulation of data on the failed restorations and not on that which may be present but are successful. It is also difficult to calculate restoration failure rates when no base line data is presented (Rykke, 1992). Despite this, it has been suggested that cross sectional studies may be a reliable predictor of restoration longevity in the practice setting if sufficient numbers of and accurate clinical histories are available for both failed and successful restorations (Mjör *et al*, 2000b; Jokstad *et al*, 1994). As the dates of placement of the restorations were known and the authors examined the restorations at known point in time the study fulfils the basic requirements of a cohort study (Crombie, 2005). Generally, it is the robustness of a study's design that is paramount when drawing conclusions from data rather than the volume or manner in which the data were collected.

Summary on cross sectional studies

The value of cross sectional data is variable and drawing parallels from such studies is dangerous. They do however present a source of data that can be valuable to the practitioner and the researcher. Despite their failings, cross sectional studies are used regularly to present restoration replacement data relating to specific populations at specific points in time (Mjör *et al* 2002; Al Negrish, 2001; Mjör *et al*, 2000a and b; Burke *et al*, 1999; Mjör 1997; Pink *et al*,

1984; MacInnis *et al*, 1991; Klausner *et al*, 1987; Qvist *et al*, 1986a and b; Klausner and Charbeneau, 1985; Boyd and Richardson, 1985; Mjör 1985). Where cross sectional studies have great relevance, is in the comparison of studies carried out at different times amongst similar patient groups. It may be possible to glean some extremely useful and clinically relevant findings especially with reference to changes and trends in material use e.g. the increasing use of composite resin in Scandinavian countries (Mjör *et al*, 2002b). Additionally, the cross sectional study can help highlight the similarities and differences between different countries and different service provision modalities e.g. independent versus state supported sectors (Mjör *et al*, 1990, Qvist *et al*, 1990b).

However, where cross sectional studies are particularly unreliable is in their “guesstimation” of restoration longevity. Significant numbers of restorations reported on in cross sectional studies do not have a reliable history or date of placement let alone whether they were replacements or first time fillings. Many cross sectional studies repeatedly rely on the accuracy of clinical records that are notoriously poor (Mjör, 2000; Mjör, 1997; Qvist *et al*, 1986a and b). Another problem of many cross sectional studies is that they ignore the existing restorations and what is likely to happen to them. Equally it is difficult to ascertain for certain cause and effect relationships e.g. if a patient attends with poor oral hygiene and failing restorations it is difficult to say for certain whether or not it is the restorations that are causing the poor oral hygiene and restoration failure or *vice versa*.

2.2.2 Cohort studies and the challenges they pose

A cohort is as a group of people who share a common experience within a defined period of time (Mausner and Kramer, 1985). Cohort studies can be retrospective or prospective. Historical or retrospective cohort studies are more common than prospective studies (Mausner and Kramer, 1985). Cohort studies follow a specific outcome on an identifiable group of patients; they can, but do not always, have a control group. The overall defining criterion of a cohort study is that time flows forward (Crombie, 2005). This occurs even if the study is retrospective, i.e. the patient cohort relates to an identifiable point in the past. Many cohort studies are labelled as longitudinal studies in the literature. A number of these are detailed in Table 2.5. Cohort studies are generally observational in nature and cohort study groups can be homogenous or heterogeneous in composition. An example of a heterogeneous cohort study in dentistry would be one that examined the life expectancy of amalgam restorations in a group of patients (Akerboom *et al*, 1993). A homogenous study would look at the life expectancy of a specific type of amalgam. Prospective cohort studies are preferred as incidence rates can be calculated and they reduce the likelihood of bias in the study. They can be used to define what is being looked at and it is more likely that important material is collected in an appropriate manner. Prospective trials set up to measure things with low incidence, need to be large and have long-term follow-ups. Cohort studies, like clinical trials, can assess clinical effectiveness. Unfortunately, cohort studies are generally a poor choice for such measurements or assessments as they often have unclear end points and, very often, lack confirmation of suitably trained and validated assessors. Problems can be reduced by using retrospective cohort studies but they themselves, as will be illustrated, have their disadvantages.

2.2.2a Prospective cohort studies

A paper by van Nieuwenhuysen *et al* (2003) exemplifies a typical prospective longitudinal cohort study. The aim of the study was to identify risk factors for the failure of extensive posterior restorations in Louvain, Belgium. During a three-year period, 926 restorations (including 89 crowns) were placed in 428 patients. The restorations were then followed for 17 years, the restorations being the cohort not the patients.

Smales' (1991) and Martin and Bader's (1997) studies provide examples of other prospective cohort studies which can also be used to highlight a number of discrepancies and reporting problems. It is important to re-iterate that a number of the challenges alluded to are not necessarily unique to the chosen example studies.

Challenges relating to operator numbers

In the van Nieuwenhuysen *et al* study (2003), the same person undertook all the restorative work and evaluations. This immediately leads to the conclusion that the results can only be operator specific and cannot be compared to another clinician or operator. Additionally, a single operator/ evaluator has the potential for a strong bias to be introduced into the study; a less than scrupulous operator or evaluator can have significant bearing on the results. While one operator is undesirable and makes the value of the results marginal, it does decrease the challenges that can be experienced when using multiple operators. From a researcher's view, it is desirable to know who the operators were, how experienced they were, or whether they were graduates or undergraduates: this information is not always readily available in cohort studies (Smales, 1991).

However, as practitioners are probably more interested in the clinical relevance of clinical techniques and practice to their own working environment the use of multiple operators would appear to be advantageous on more than one count.

Challenges relating to patient selection

Van Niwenhuysen's study (2003) used a highly selected patient group with recruitment aimed at a cohort with a high level of dental awareness and relatively high socio-economic status; it is debatable whether dental hospital patients are representative of a local patient population.

Challenges relating to changes in dental materials

It is not unusual for dental materials to be suddenly withdrawn from the market or be replaced by "superior" modifications of the original and this can be a problem in a study that is running over an extended period. In van Niwenhuysen's study (2003) two different amalgam and three different composite materials were used. Additionally, three different base materials were employed and although not a major problem in a Cohort study, it could be in a clinical trial.

Smales' (1991) study looked at cuspal coverage and non-cuspal coverage restorations placed in patients attending the Adelaide Dental Hospital. High copper or more conventional amalgam alloy restorations were placed in premolar and molar teeth with and without the use of pins. One conventional alloy and three high copper alloys were used throughout the study and there is no mention in the paper as to how amalgam types affected the results. It has to be remembered that generic material grouping e.g. all high copper alloys do not confer similar mechanical or physical properties.

Challenges relating to clinical technique

It is not uncommon to find inconsistency with respect to clinical techniques in cohort studies with little reference being made to the standardisation in clinical procedure e.g. whether one type of threaded pin was used or indeed the overall composition of this pin-retained group of restorations (Smales, 1991).

Challenge relating to tooth variables

Despite the large numbers of restorations in the van Niwenhuysen *et al* study (2003), 926 in 428 patients (722 amalgams, 115 composites and 89 crowns). It is clear that a number of tooth variables are presented for analysis and that is before further variables such as restoration size (the authors split them into partial and complete restorations), gender or age are investigated. Van Niwenhuysen's study (2003) examined restorations placed in premolars and molars and mandibular and maxillary teeth. Some of the restorations utilised auxiliary retention measures (624) others did not. There were differences in tooth vitality with 60% of the teeth being treated endodontically.

The above highlights the relative complexities of many cohort studies and the difficulties that can be encountered when trying to examine specific objectives over time. Very often the end result is small numbers of restorations in multiple subgroups despite the study apparently looking at a significant number of restorations e.g. amalgam restorations in maxillary premolars.

Challenge relating to drop-out

There is a general problem of drop out in cohort studies. Van Niwenhuysen *et al*'s (2003) drop out was 41% in the first period (926 to 536 restorations), 38% in the

second (a reduction from 526 to 323 restorations), 25% in the third (326 to 243 restorations), 15% in the fourth (243 to 154 restorations), 2% in the fourth (154 to 151 restorations) and 0% in the fifth. The overall drop out was 84%. Of the 151 restorations evaluated over the 17 years 48% were functioning well and 28% had failed.

Martin and Bader (1997) previously highlighted the problem caused by drop-out and suggested they should be excluded from any analyses. Van Niwenhuysen *et al* (2003) compared the drop out groups in their study cohort and showed little differences with respect to gender but a slight difference in age (41 compared to 47 years of age) but importantly highlighted how such analyses can show that the remaining cohort is representative of the original sample overall.

Challenges relating to data manipulation

Martin and Bader's (1997) paper on 4,735 posterior restorations, followed for five years, in an insurance scheme used by 74 dentists shows a number of data reporting discrepancies. This paper reported rounded percentages rather than actual numeric values and consequently generated reporting errors of around one percentage point. These authors also published a table detailing treatment outcome over five years based on restoration type and three Kaplan-Meier survival curves that show significant differences between the survivability of crowns and amalgam restorations over the time period. Unfortunately, they did not provide the reader with enough raw data to substantiate these findings. Such findings are not unusual when reporting cohort studies and clinical trials. It is realised that this may not actually be down to the authors but occasionally consequent of the review process during publication and the difficulties imposed by journal editors who often

curtail the amount of data published in their journal. However, without raw data, studies cannot be combined and this can result in otherwise good research being excluded from a systematic review (Smales and Hawthorne, 1996; Smales, 1991; Bentley and Drake, 1986). In order to overcome some of the difficulties that can be experienced in the reporting, interpretation and understanding of clinical research the CONSORT statement was published in the mid- 1990s. This statement has been adopted by a number of editorial groups and has been shown to help in the quality of clinical research reporting (Moher *et al*, 2001). Although the checklist and associated flowchart relating to 22 items pertains to randomised controlled trials there is no reason to doubt its impact should it be applied to the reporting of all clinical studies and allow readers to fully understand a studies conduct and assess the validity of the published data.

Challenges relating to data comparison

In their discussion, Van Niwenhuysen *et al* (2003) mention the difficulties of comparing their data set with those of other longitudinal studies that assess the behaviour of amalgam and crowns (Smales and Hawthorn, 1997; Martin and Bader, 1997; Smales, 1991; Bentley and Drake, 1986). Never the less these authors still made a comparison with other studies and suggested that their results are superior despite the fact that they compared their results with a studies of a completely different design and different patient cohorts; the Bentley and Drake (1997) and Smales and Hawthorne (1997) studies were retrospective. Although the Martin and Bader (1997) and Smales (1991) studies were prospective they used different clinical techniques to the van Niwenhuysen study. Additionally, the Martin and Bader's (1997) study was carried out in the independent sector with the

Smales (1991) study being completed in a dental hospital.

2.2.2b Retrospective cohort studies

Like prospective cohort studies, retrospective cohort studies present with their challenges.

Challenges relating to data collection

A typical retrospective cohort study is represented by a publication that examined the performance of extensive posterior amalgam restorations and crowns in private dental practice in Australia (Smales and Hawthorne; 1996). This study examined the case notes of 300 patients that fulfilled their acceptance criteria (continuous attendance at one of the three identified private dental practices for at least fifteen years). The number and location of the restorations in the selected patients were scrutinised and evaluated: the records detailing the date of placement of the restoration and any subsequent failure, repair or replacement. It is clear that accurate data recording in the clinical record is the corner stone of this research modality: lax recording significantly hampering data collection and analysis (Levin, 2006).

Although Smales and Hawthorne (1996) mention that 160 extensive amalgam restorations, 96 full gold crowns and 174 ceramo-metal crowns, were examined they do not indicate what proportion of restorations were attributed to a particular operator or dental practice. The authors censored their failure data into that which could be attributed to the material and that which could be attributed to operative technique. They also detailed other: replacements for endodontic purposes, periodontal or other dental reasons e.g. replacing an amalgam with a

crown as the tooth became a bridge retainer. Unfortunately, no information relating to specific material type was provided nor did the authors indicate in the paper as to where and when failures occurred. Therefore, this paper only becomes of relevance to restorations that have survived intact over the period of examination. It is only possible to say with certainty that of the original X% Y % survived in this cohort; the values of the results are limited with reference to other populations, other studies or other cohorts. Although their 5, 10 and 15-year survival rates suggested that crowns appeared to be more successful as a treatment modality these findings need to be examined carefully, e.g. we cannot say whether the teeth without the crowns were unsuitable for crowning in the first place; perhaps they had a poor prognosis and were deemed unsuitable for crowning. Such a decision would be removed in a well-conducted clinical trial as treatments would be allocated randomly and the results measured accordingly.

In order to eliminate some of the challenges in determining restoration longevity from the retrospective evaluation of clinical records it is possible to identify patients and or restorations and follow their clinical history forwards to a specified clinical examination. Bentley and Drake (1986) undertook this approach in their evaluation of restoration longevity in an American Dental School. From a group of 3000 patients 86 fulfilled their acceptance criteria (ten years continual attendance at the clinic with the patients having only received routine hygiene recall). That so few of the 3000 patients were included (2.86%) suggests that the sample may have been atypical. However, data on 1,207 restorations was studied. Previous failures and replacements (ascertained through examination of the clinical record) counted alongside the calibrated evaluations of the remaining restorations by the authors. The presented data could be determined as being

indicative of clinical treatment provided in the unit and verifiable through the validated examinations. As the examinations and recordings considered both successes and failures, true-life table analyses relevant to the cohort studied were presented. As is common in retrospective cohort studies the authors identified that the previous recorded reason for failure could not always be verified. The actual reason for 99 of the 351 failures was established; this obviously skews the data presented. This illustrates the difficulty of determining failure retrospectively from clinical case notes that have not been set up to record the desired data.

Summary on cohort studies

Despite the problems alluded to in both prospective and retrospective cohort studies there is much valuable information that can be gleaned from and highlighted by them.

There are examples of good cohort studies in the literature that highlight good research practice e.g. Akerboom *et al* (1993). This study detailed the number of operators (3), the number of restorations (1, 544), the number of restorations each participant received and the evaluation criteria. While, Akerboom *et al* (1993) examined amalgam failure in general terms they also provided the reader with a detailed and accurate recording of the types of failure and its subsequent statistical analysis with respect to tooth, size, alloy and operator. This paper also attempted to explain its findings with respect to the particular group of patients seen, treated and evaluated rather than make over-generalisations.

It appears that longitudinal prospective cohort studies can be used to collect data on restoration survival. However these studies need to be of sufficient size

and can pose challenges which may be neither practical nor workable e.g. a cumulative 1% failure rate in a cohort of 1000 restorations would only reveal failure data on approximately 94 restorations after ten years.

Consequential to the challenges highlighted above it is not surprising that many cohort studies fail to reach the exacting standards required of a systematic review and to that end the information gleaned from many cohort studies must be deemed of limited value when it comes to measuring restoration longevity.

One big advantage of cohort studies is the relative ease in which data can be collected and analysed when the desired data has been collected in a standardised and accurate manner. This standardisation in the collection of clinical material is what makes a longitudinal prospective cohort studies appealing. It is also what makes it difficult in the long term. Results, even with problems such as fall out, can be collected and trends observed from data analysis; the ability to define and report treatment outcomes becoming a particularly appealing factor. This is particularly pertinent if the study is designed in an appropriate way with a suitably representative population and appropriate number of operators from a specific and generic environment. If this can be established it is quite probable that the results may have direct bearing to clinical practice and give an indication of what can be achieved. This information could then be used as a baseline of comparison and the findings then possibly form the basis for other clinical trials.

2.2.3 Clinical trials and the challenges they pose

The randomised controlled trial (RCT) is a type of clinical trial, or scientific procedure used in the efficacy of medicines and medicinal procedures⁶. It is widely considered the most reliable way to accumulate scientific evidence, as it is the best design for eliminating the variety of biases that regularly compromise the validity of medical research. For clarity and the purposes of this report, it is important to distinguish between a clinical study and a clinical trial. In this section a clinical trial refers to an attempt to assess the merits (or otherwise) of dental materials in a clinical experiment or the evaluation of clinical procedures.

There are many challenges that can be experienced during the design and running of clinical trials.

The challenges posed by effect modifiers

Effect modifiers are often very difficult to control within a clinical trial. Chadwick *et al* (2001) identified a number of these factors and divided them into two broad categories: objective and subjective. The objective group contained three subgroups. The first related to patient factors (exposure to fluoride, caries status, the patient's health, the patient's parafunctional activity, age, xerostomia, diet and socio-economic factors). The second to tooth factors (tooth location, tooth size, tooth type, cavity design, cavity type, type of dentition, occlusal load placed on a restoration, tooth quality). The third related to clinical operating factors (type of material used, physical properties of the material, quality of finish of the restoration, moisture control, use of local anaesthetic, clinical expertise, clinical training). The subjective factors included influences such as payment structure,

⁶ (http://en.wikipedia.org/wiki/randomized_controlled_trial, 16/05/2006 19.34hrs).

the clinical setting, the country in which the work is carried out, clinical diagnostic, treatments and maintenance philosophy and patient preferences. Effect modifiers are evident in other types of clinical research and these have been dealt with and mentioned elsewhere.

The challenge of bias

Bias is a systematic tendency to overestimate or underestimate a population parameter (Bulman and Osborn, 1989). Bias reduces the validity of the results obtained in a clinical trial (Faragher and Marguerie, 1998). It can be introduced into a clinical trial from various sources; the patient, the operator, the researcher, the statistician. It is argued that if any of the aforementioned groups in a clinical trial are aware that they are undergoing treatment (or evaluating treatment) or acting as (or assessing) a control group in an experiment the opportunity for bias in reporting exists. Cunningham *et al* (1990) give an example where this may not be the case. They stated

“... a restoration was deemed to have failed if, in the opinion of the clinician, it required replacement...”

The challenge of blinding

Blinding in clinical trials helps to ensure that placebo effects and bias in the interpretation of handling of a particular patient group is minimised (Faragher and Marguerie, 1998). Blinded clinical trials and in particular double or even triple blinded trials where operator, patient and evaluator are unaware of the particular “product” used are, arguably, the gold standard. Unfortunately, it is not always possible to conduct such clinical trials in dentistry. Experienced operators can

quite easily distinguish between different material types, on occasion, (Bates and Douglas, 1980). Again, this can introduce bias into the evaluations unless efforts are taken to make sure that the evaluations are truly objective

The challenge of randomisation

Byar (1976) has defined randomization as -

“... a procedure for assigning treatments to patients in such a way that all possible assignments of treatments to patients are equally likely within the constraints of the experimental design.”

Prospective, randomised, blinded and controlled clinical trials offer the researcher the greatest understanding in an examined field if the trial is carried out properly (Duke, 1992). Prospective trials enable the instigation of randomisation from the outset of a trial. Randomisation guards against selection bias between participants and operators or evaluators. It creates groups that are comparable in all factors that may influence the prognosis and it gives validity to the statistical treatment of the data (Mausner and Kramer, 1985).

When carrying out a clinical trial to compare dental materials it is relatively easy to have randomisation protocols built into the trial in order to reduce patient variables. However, while the allocation of procedures can be randomised doubt exists in the ethics of such protocols in dentistry when aesthetics is an issue for the patient. Arguably, it is unethical to “force” a patient to accept something that when allocated to they object e.g. an amalgam restoration being placed in what the patient finds as a visually unacceptable location. This has particular relevance in dentistry when the subsequent replacement of a restoration will increase the morbidity of the tooth. Without randomisation procedures truly unbiased results

cannot be obtained, arguably it is unacceptable not to randomise treatment methodology in clinical trials.

Challenges associated with the operator

Despite it being preferable to involve a number of clinicians in a clinical trial the clinical evaluations of dental materials based on the operative efforts of individual clinicians have been reported (Berry *et al*, 1995). It is important to consider exactly how many operators are required for the results of a clinical trial to be both realistic and workable. The problem with a small group of operators e.g. three (Cunningham *et al*; 1990) is the possibility that one operator can significantly skew the results; one operator could produce a significantly poorer or a significantly higher quality of work or evaluations within the trial and data needs to be analysed carefully to safe guard against this. It is also important that operators be matched with respect to skill and training relevant to the trial being conducted. This is possible with a small group but becomes increasingly more difficult as the number of operators increases. Sub-group analysis, therefore, becomes an important part of data interpretation in clinical trials. The effect of operator needs to be evaluated and the integrity of the results ensured by such.

The challenge of environment

The majority of published controlled clinical trials have been carried out in dental schools or hospital environments. The value of hospital based clinical trials has been debated (Letzel *et al*, 1978, Mahler and Adey, 1977; Eames and McNamara, 1976; Duperon *et al*, 1970; Mahler *et al*, 1970) and it has been said that they have little value to the average general dental practitioner; never the less it should be a

guide to it (Osborne, 2006). Bates and Douglas suggested in 1980 that in order to obtain realistic and clinically relevant results for general dental practice clinical studies should be carried out in general dental practice. One significant drawback of carrying out hospital-based trials is that local exclusion criteria often make the population cohort taking part self-selective e.g. an ability to attend for regular follow up, availability of particular dates etc. Such recruitment procedures can make the generalisability of results questionable.

Significant steps are taken in clinical trials to make sure that a particular group of patients do not exert undue biases. It being ideal if the trial is conducted on a sample truly representative of the population. This can be difficult in clinical trials as the very nature of patient recruitment can theoretically bias the results by inadvertent behavioural changes resulting from the process of simply being recruited into the trial. By their nature, clinical trials are often intensive on patient time, require extensive follow up and in order to reduce the problems of drop out often require extremely motivated patients. It is arguable if such a cohort is ever representative of the general dental population. It is not unusual to find that there may a particular skew to the patients recruited into a clinical trial e.g. the unemployed, student bodies, staff and patients based in a particular environment, highly motivated, eager to please dental hospital patients or even inmates. However, as mentioned earlier, once enrolled it is fundamentally important that patients participating in a clinical trial are randomised with respect to the treatments they receive if bias in the trial is to be avoided. Good clinical trials detail how the randomisation was achieved.

It is desirable that patients are treated in a single or at the very least similar and comparable environments with similar facilities and equipment if undue bias is

to be minimised. Clinical facilities need to be replicated in all places of work or assessment with things like standardised lighting and standardised instrumentation being provided. This is something which can be easily overlooked, or at the worst taken for granted. Cunningham *et al* (1990) are not explicit in their paper as to where the patients in their trial were treated. They state that two of the three operators were primarily general dental practitioners but the paper does not mention whether the patients were treated in general practice or the hospital environment.

Equally, if assessments are being made on models or photographs it is not only important that the models or photographs are produced in similar conditions but that they are also evaluated under similar conditions i.e. with or without magnification and again in standardised lighting.

The challenge posed by multiple subgroups

It is easy to introduce multiple variables within subgroups in clinical trials.

Subgroups complicate clinical trials and they can quite easily discredit the value of the information gleaned from the trial. This is highlighted in Cunningham *et al*'s (1990) paper that evaluated five materials; three posterior composites and two amalgams. Overall, the trial was concerned in the differences between two major classes of material (amalgam and composite resin). However, Cunningham *et al* (1990) used different materials within their main groups of materials e.g. two light-cured and one chemically cured posterior composite. The composite resins were different and there is the possibility that one particular product could have superior or inferior physical properties due to variation in resin and filler content. A similar problem can be seen in the amalgam group where the potential for differences in

alloy composition or methods of manufacture existed; the complexity and potential for variation being illustrated in Cunningham *et al*'s (1990) trial. Undoubtedly, an ability to make comparisons between amalgam and composite existed. However, internal differences within the materials could affect the results depicted by the main group e.g. all the failures may fall into one group or produce similar failures; while these intra-material comparisons are interesting they are, once again arguably, superfluous if the trial is not set up to measure such in the analysis, (Crombie, 2005). When using different materials in a clinical trial it is important to remember that the distribution of failure may not be the same within different classes of the same material. It is important that the study design accounts for this and where necessary makes provision for sub group analysis by recruiting sufficient numbers into the research. It has also to be remembered that if you only use one type of composite and compare with one type of amalgam then you can only extrapolate your results to this amalgam or this composite.

Challenge relating to operative procedure and material variability

When multiple operators are used in clinical trials it is important to know that similar methods and operative procedures were employed. Examples of some very important questions that need to be asked in dental restoration based trials would be: Were they all placed under rubber dam? Were they all lined with the same base material? Were they all placed in the same manner? Were they all finished in the same way? Were they all completed within a specified time? As noted earlier, Cunningham *et al* (1990) use different dental materials in their trial. There were also differences with regard to material placement. One of the posterior composites (P30TM) was placed with an occlusal margin bevel where the

others were not (Occlusin™ and Clearfil™). Once again, a source of potential conflict is introduced into the trial.

Clinical trials which evaluate dental materials over an extended period of time have their own peculiarities. Dental materials do not always have what can be considered an infinite shelf life and a material's shelf life can dictate a trial's length. Equally, it is not unusual to find that a dental company's wish to improve their materials may affect a material property; if one is not careful one can easily end up using a similar but different material on a later date in the trial. While the chances of this happening during a clinical trial may seem remote, it can happen if steps are not taken to prevent such an occurrence. Darvel (1978) suggested that a single batch of material is used throughout clinical trials and that the restorations are placed in a sequential manner by the operators i.e. one operator placing all his restorations before the next does theirs etc. The rationale behind this is the ability to determine if there is any degradation in material properties. While material degradation may be discernable by such a protocol, it does not address the fact that there is a chance that the last restorations placed by an operator could be of a higher or lesser quality than the first ones placed. If operators place the restorations in batches it can be difficult to determine if it is the material or the operator that is at fault. It is also possible that minor material degradations will influence the results of the trial e.g. because of the evaporation of a volatile component from a material.

The challenge associated with standardising measuring and recording

Whatever method is used to evaluate the results in a clinical trial it is important that validity and reliability be assured in the recording and evaluation processes.

Validity is required to ensure that there is a degree of certainty in what is attempting to be measured is actually being measured. Valid measurements have degrees of sensitivity and specificity. In human disease, sensitivity identifies the disease being looked for; specificity determines cases that do not have a disease. Reliability is the consistency with which a measurement or assessment is made. The validity of the observations or measurements needs to show a degree of robustness in a clinical trial. This robustness should not only be documented but substantiated where necessary (Crombie, 2005).

When new or previously unreported evaluation methods are being reported there should also be some form of guarantee that the methods used to evaluate whatever is being evaluated is valid, acceptable and reproducible and wherever possible objective; subjective opinion should be kept to the minimum.

Cunningham *et al* (1990) present a situation in their paper that exemplifies a need for such clarification and justification. They use a four point scale to score for colour match. However, there is a considerable degree of scope for overlap and error of recording e.g. when does a slight mismatch become an obvious mismatch? Surely, if a mismatch is identifiable then it is obvious. Additionally, should an evaluator score up or down if he is in doubt? It is important that evaluations be founded on standardised and trained methodology that is suitably validated to confirm that an evaluation is wherever possible correct, that undue bias has not crept into the study and individual subjectivities minimised. This would be particularly relevant in studies that examined restoration replacement where singularly identified and respected evaluation criteria are used.

The challenge of obtaining standard cavity types

Clinical trials designed to assess the merits or otherwise of dental materials very often use a split mouth design. This helps to reduce (but not eliminate) a number of patient associated factors and reduces variation between patient cohorts. Split mouth designs are not suitable for clinical trials that may affect the mouth as a whole e.g. the use of two dentifrices in one mouth at the same time and then attempting to evaluate its effect on only one side of the mouth. Ideally, contra-lateral teeth in the same jaw of a patient should be used for double blind clinical trials (Bates and Douglas, 1980). This paired design can be used to help standardise a number of confounding variables found in the trial and while caries is generally considered a symmetrical disease, this is not always the case. It is not easy to recruit patients into randomised, paired and double blinded clinical trials; this difficulty was noted even when caries rates were relatively high in the UK (Bates and Douglas, 1980). Equally, due to the nature of dental caries and the infinite way that it presents it is difficult to limit the recruitment of patients into clinical trials with specific cavity types unless it is carried out over a long time or is multi-centre in design. The challenges posed by multi centre designs has been dealt with in the problem areas of evaluation, the use of multiple operators and the difficulties experienced between different patient cohorts that will naturally occur between centres. Conducting trials over a long period can accentuate the difficulties of patient retention throughout the period of the trial.

This challenge, recruiting cavities of a particular type, is quite real in clinical dental trials. This challenge is again highlighted by the Cunningham *et al* trial (1990). In this trial 605 restorations were placed; 83 occlusals, 204 two surface restorations (140 MOs and 164 DOs) and 122 three surface restorations (MODs).

The proportion of materials used in each of the five groups was roughly similar in that around one fifth of the total number of restorations belonged to each material. However closer examination of the data reveals dissimilarities in the recruitment of particular cavity types to particular materials. The proportion of occlusal restorations ranged from 12 to 26 percent of the total for each of the materials, 16 to 22 percent for the mesio-occlusal restorations and 13 to 22 percent in the three surface group. These apparent “anomalies” need to be analysed to ensure that they were no modifying effects on the results. Cunningham *et al* (1990) fail to provide the reader with any indication on restoration location and the reader cannot work out whether they were in the mandible or maxilla or what proportion of the restorations was in premolars or molars. It is not possible to work out for example if all the occlusal restorations were only in premolars. The larger a trial is the less likelihood that such problems could bias the results. However as seen here, even apparently large numbers of restorations can relate to relatively few specific teeth or cavity types.

The challenge of follow up

As with cohort studies, clinical trials can lose significant numbers of patients to follow up and the number of patients or restorations lost to follow should be documented. Cunningham *et al* (1990) evaluated nearly 85% of the originally placed; this is an excellent return after three years. Other trials show less favourable follow up data; Barr-Agholme *et al* (1991) 12% after two years; Hamilton *et al* (1983) less than 60% after 5 and less than 30% after ten; Bates and Douglas (1980) 86% after two years. It has been suggested that follow up rates with less than 90% of the original sample can generate significant problems in

data analysis as the true condition of the missing restorations cannot be determined –they could all be perfect or equally all total failures (Mausner and Kramer, 1985).

Challenges related to identifying and measuring failure

Cunningham *et al* (1990) highlight some of the challenges when measuring or evaluating failure. It being reported that 35 (6.87%) of 509 restorations failed over the examination period; the failures being either mechanical or biological.

Expanding on this, ten of the failures could be attributed to non-material faults (pain, caries, periodontal reasons) while the rest were attributed to material faults (tooth fracture, filling fracture and loss of the restoration. A total failure rate of only 4.9% is recorded for the so-called material failures; assessing failure on such small numbers of restorations needs justification, clarification and an evaluation of the certainty, especially when the results are often extrapolated to predict failure as a whole. It is also difficult to classify restoration failure as being purely based on material failure as restoration placement is highly reliant on operator skill and patient factors as well (Burke, 2006a; Burke 2006b). This challenge of insufficient numbers becomes particularly relevant when looking at intra-material differences in clinical trials that are either incorrectly designed or too short. For example in Cunningham *et al's* (1990) study when looking at the incidence of fracture within the three posterior composites it is seen that one group had eight failures another two and the last six. It is doubtful if meaningful comparisons can be made from such data.

The failure rate reported by Cunningham *et al* (1990) is slightly lower than the incidence reported in other clinical trials that varies from 8.4 to 11%

(Sturdevant *et al*, 1988; Robinson *et al*, 1988; Wilson *et al*, 1986). Robinson *et al* (1988) also reported no amalgam failures in his study whereas Cunningham *et al* (1990) reported an amalgam failure rate of nearly 80%. One way of overcoming such difficulty with respect to having enough meaningful numbers on material failure is to combine the results from a number of similar studies. This is a process known as meta-analysis. Meta-analysis is a process that generally involves the extraction of data from a systematic review, subsequent computing of summary statistics for each trial, weighting of these statistical values and then averaging the summary statistics to produce an overall effect size with confidence intervals. It is important to note that meta-analyses and systematic reviews are not synonymous terms as not all systematic reviews will conduct a statistical meta-analysis and there are times when this is neither possible nor indeed practical. In such cases, a qualitative comparison between studies will be undertaken.

The challenge of time

Clinical trials are often constrained by time; companies can be impatient for results, researchers are keen to show their findings, long clinical trials are expensive to run, funding may be limited and drop out rates increase with time. Consequently, many clinical trials are of a short nature and few extend beyond a period of three years. As mentioned, this can lead to difficulty when absolute or long-term failure rate is being evaluated. Despite this, short-term clinical trials can be of great benefit in the evaluation and determination of the clinical suitability of dental materials and it is not unrealistic to extrapolate significant failure rates from relatively short-term trials to a more clinically relevant period of time (Berry *et al*, 1995). Few disagree with the principle that short-term clinical trials are needed as

they can identify early weakness, failure or problems with specific materials. It is probably also true to say that, the use of unsound materials is better determined on ethical grounds in the laboratory than in the field. Short-term clinical trials however do not have much value when no difference between test groups occurs at the end of the evaluation period (Berry *et al*, 1995). They also have limited value when utilising certain parameters to determine future failure e.g. marginal degradation (Mahler and Marantz, 1979). It is important that when clinical data are presented at an early stage in the study that the patient cohort is followed and subsequent data published. Accepting the results of relatively short-term data at face value is fraught with danger. A good example of this is shown in the results from a study carried out in Newcastle, England that assessed the performance of amalgam and glass ionomer cement in deciduous teeth (Welbury *et al*, 1991). Welbury *et al* (1991) reported a significant difference in median survival time for amalgam over glass ionomer cement while an earlier report from Walls *et al* (1988) indicated that there was no difference between the materials.

Summary on clinical trials

There are few well-conducted randomised controlled trials in the dental literature that satisfy the rigours of a systematic review; there are even less on restoration replacement. The clinical trial is a study type that has, perhaps, the greatest scope for producing clinically relevant results. Perhaps it is not so surprising that there are so few double blind, randomised, controlled clinical trials in dentistry as it is clear that they are subject to a number of ordeals and tribulations and that they can be very difficult to conduct unless a considerable degree of planning goes into them. It is clear that the major parameters of clinical trial design, conduct, analysis

and interpretation are of fundamental importance in the evaluation of clinical trials and no less so than in the conduct and evaluation of other types of clinical studies. Ironically, the very things that make a randomised controlled trial desirable appear to make them difficult to conduct; it being noted that the many variables that exist in clinical trials preclude them from being compared to each other (Mowafy *et al*, 1994).

When evaluating clinical research in dentistry there is a need to be aware that it is not only experimental design or material faults that create difficulty but that there are additional problems associated with the operator, the patient and the clinical environment where work was carried out. The assessment methods used to evaluate restorations are often varied and dependent on the diagnostic criteria used and how they are interpreted and applied by the assessor.

It has to be borne in mind that the controlled clinical trial is exactly that ... “a **controlled clinical trial**” and that its findings are not necessarily directly applicable to the average dental practice environment. Consequently, it is necessary to rely on other sources with less robust data from studies with less robust research methodologies to come up with answers to certain questions. Additionally, the usefulness of data that is generated by rigorous and standardised methodologies that rely heavily on the individual skills of the dentists taking part in the study has to be considered. It is important not to dismiss readily the value of other clinically based research such as cross sectional studies.

2.2.4 Problems associated with different types of clinical study⁷: An evaluation of the literature by limiting the field

As shown above, different clinical studies have varying degrees of clinical and research robustness. Clinical studies are used to determine whether one clinical treatment method is better than another. The focus of clinical studies is invariably the outcome.

In addition, to grouping studies with respect to their design e.g. cross sectional studies, longitudinal studies, prospective and retrospective studies clinical studies can be observational or experimental. Such divisions are unfortunately rather generic and results in large groups of similar studies that cannot be compared readily. However, the systematic review can be used to identify studies by both design and outcome measure with studies being grouped from weakest to strongest (reports from expert committees, case studies, retrospective case series, prospective case series, retrospective studies with concurrent controls, prospective studies with historical controls, prospective studies with concurrent controls, clinical trials other than randomised controlled trials, randomised controlled trials) and also with respect to the strength of the outcome measure (subjective opinion, criterion based decision making, criterion and training based decision making but lacking calibration, calibrated training and criterion based decision making, failure without intervention). This cross tabulating of research by type of study and outcome measure is useful and represents a process carried out in good systematic reviews (Chadwick *et al*, 1999).

As mentioned, the gold standard in clinical research is the prospective randomised controlled and blinded clinical trial. In restorative dentistry there are a

⁷ The figures in brackets represent the study design and outcome measure used to classify the paper evaluated. These refer to the descriptor of the columns and rows seen in Table 2.7.

limited number of studies that manage to fulfil the exacting standards set out in a systematic review (Chadwick *et al*, 1999; Downer *et al*, 1999; Worthington *et al*, 1997). Chadwick *et al* (2001) only identified fourteen randomised controlled trials in their systematic review.

Clinical trials have their advantages and disadvantages. Very often, the generalisability (i.e. the applicability to every day general dental practice) of the study is limited because of one or a number of factors. Some of these problems have been highlighted already but others with referenced papers fitting specific categories are highlighted in Table 2.1.

With specific reference to restoration longevity, a systematic review of the literature carried out by Chadwick *et al* (2001) revealed only 253 of 5,675 clinical studies concerned with the longevity of dental restorations to be worthy of inclusion in their review. The selected 253 was reduced to 195 when link papers were removed from the report. As is customary with systematic reviews, Chadwick *et al*'s (2001) report stated the basic criteria that had to be fulfilled for acceptance into the review. With respect to determining restoration longevity the authors highlighted some of the major problems encountered in their review;

- drop out rates needed were not always clearly stated,
- it was not always possible to deduce from the data what the actual failure rates at each time period were,
- baseline data was often missing,
- when USPHS criteria were applied it was not always possible to determine whether single or multiple faults were being reported,
- randomisation was lacking in some trials,
- the use of inappropriate analyses e.g. reporting median survival time

when parametric modelling had not been appropriately applied,

- the lack of detail in some papers with respect to effect modifiers e.g. clinical procedures involved, population characteristics, control groups, the study environment, the objectivity of the criteria used to identify failure, operator characteristics, payment system or sponsorship.

A summary of their included studies based on outcome measure and study design is shown in Table 2.6. As will be seen later a number of the studies that they rejected in their review reveal valuable information when it comes to trying to answer the question “How long does a restoration last?” In order to highlight some of the challenges that may need to be overcome in clinical studies an example from each of the different study types based on study design and outcome measure have been used. It is, however, worth pointing out that despite the apparent bleak picture of clinical research portrayed in the following sections many clinical studies reveal unique insights into differing fields, provide knowledge, and generate questions that fuels the researchers desire to answer, prove or refute. The drive to answer research questions while making sure that the correct research tool is used is probably the ultimate goal for many clinical researchers.

It has to be remembered that clinical studies are not infallible; it is the extent and degree of the flaws that determines the robustness and reliability of the data and its interpretation (Crombie, 2005). The next chapter illustrates these points by grouping the literature together by limiting their field.

Prospective case series using restoration replacement criteria without training (3, 2)

Barnes *et al*, 1991.

This eight year study carried out in the USA was designed to evaluate Class I and Class II composite (Ful Fil™) restorations. Thirty three restorations of which 25 were Class I and 8 Class II were placed in twelve patients of an indeterminate age. Of the 33 restorations, 23 were placed in molars and 10 in premolars. After five years, 32 restorations were still in service with one failure noted. At eight years, 30 restorations were available for evaluation and two failures recorded. The paper details the clinical techniques used by the clinicians. Unfortunately, it is not possible to tell how many clinicians were involved in the research. As can be seen the major problem with this trial is that an indeterminate number of clinicians were used and that the university environment in which these restorations were placed could limit the studies relevance to general dental practice. There are also the problems associated with lack of basic data and that the failures were small in number; this limits comparison. Equally important is that the failures were not validated or corroborated.

Prospective case series using restoration replacement with training (3, 3)

Andersson-Wenckert *et al*, 1997.

This two-year Swedish study, carried out in 5 dental public health clinics, was designed to assess the performance of a compomer material (Dyract™) in Black's Class II cavities in deciduous molars. The study was based on 159 Class II restorations (144 micro-cavities and 15 conventional) placed in 79 children ranging from five to twelve years of age. This study used a modified version of the USPHS

criteria to assess the restorations placed by six dentists. The clinical techniques employed were stated and included; calcium hydroxide lining in deep cavities, the use of a priming agent and the use of a stainless steel matrix band with saliva ejector and cotton wool rolls for isolation. At 24 months, 20 of the 159 restorations had failed (5 due to caries, 7 due to loss, 3 due to a combination of caries and loss, 1 due to fracture, 3 due to marginal adaptation problems and 1 due to an unknown cause) and a further 25 teeth having been exfoliated or extracted. The failure rates between the different operators ranged from 12 to 35%. The main problem with this study is its relative shortness and the small degree of failure, especially when they are analysed by individual operator. It does however show quite nicely the sensitive nature of restoration placement technique and the fact that operator ability plays a part in restoration success rate; the authors stating that the failure rate reported was significantly higher than reported in more closely controlled studies.

Prospective case series using restoration replacement using validated outcome criteria (3, 4)

Holan et al, 1996.

This two and a half year American study carried out in a hospital environment was designed to compare Black's Class II composite restorations with Black's Class II amalgam-composite restorations in deciduous molars. The study used a modified version of the USPHS criteria and utilised two or three assessors. It is not possible from the paper to determine the number of operators who took part in the study. A total of 42 restorations were placed in 22 maxillary and 17 mandibular molars in eighteen children between the ages of six and a half to twelve years old

(12 boys, 6 girls). The clinical placement techniques for the restorations are described in detail; use of local analgesia being stated, rubber dam and calcium hydroxide linings were used. The composite restorations utilised an enamel bonding system and the amalgam restorations a dentine bonding system. Three groups were evaluated in the study; amalgam and composite with bonding, amalgam and composite without bonding and a composite only group. No failures were reported over the examination period: twelve of fourteen, sixteen of seventeen and eleven of eleven being evaluated for each group, respectively. The study lacks generalisability to the practice environment as it was carried out in a hospital. The shortness of the study and the unusual combination of material use also affect the generalisability of the findings; in this instance the researchers postulating that an amalgam lining can prevent micro-leakage under composite restorations. As well as the small numbers reported in this study it is also arguable that placing a restoration in the deciduous molar at age twelve or in a tooth that is about to be exfoliated has limited long term value.

Prospective case series using restoration replacement and measuring “true” restoration failure (3, 5)

Jordan *et al*, 1993.

This two year University / Hospital study carried out in Canada by a single clinician and two (unconfirmed) trained and calibrated evaluators was designed to evaluate the clinical performance of a photo-adhesive system in non-retentive Black’s Class V cavities. Forty-two patients between 46 and 69 years of age had 95 restorations placed under controlled clinical conditions that included rubber dam placement and enamel bevelling. At two years, 83 of the original restorations were evaluated

of which two had failed. The study lacks generalisability in that it was not carried out in a general practice environment and used rubber dam which is felt to be an unusual practice for Black's Class V cavities in General Dental Practice. Examiner agreement in this study was quoted at 85% for the two assessors which increased to 100% when a third arbitrated when dispute occurred; the arbitrator being the original operator.

Retrospective study with concurrent controls restoration using replacement criteria without training (4, 2)

Wilkie *et al*, 1993.

This two-year Australian university study carried out by two operators and two trained and calibrated evaluators (the operators) was designed to compare the survival of cermet, composite and amalgam restorations used in Black's Class II and tunnel restorations. The study was carried out with well-defined clinical techniques (local analgesia, rubber dam etc) on 26 adults of indeterminate age. Eighty-six restorations (44 Class II and 42 tunnel preparations) were placed in 35 permanent premolars and 51 permanent molars; 16 were amalgams, 42 were cermets and 28 composites. Forty-seven restorations of the original 86 restorations were evaluated after two years; 31 being Class II's and 16 tunnel preparations. No failures were reported in the amalgam group. The composite group showed a 91% cumulative survival rate and the cermet group a 45% cumulative survival rate at two years. Although the generalisability of the results cannot be applied to general dental practice the study does highlight that the cermet restoration tested was unsuitable as a permanent restorative material for permanent teeth. This paper highlights the value of well-controlled studies to

identify failure over a short period. It is, however, arguable as to whether or not operators should act as their own assessors.

Prospective study with historical controls using replacement criteria without training (5, 2)

Scheer, 1975.

This British study carried out in a university environment over three years by an un-stated number of operators or evaluators was designed to evaluate the success of fractured incisors restored with the acid-etch technique. Ninety-two children (36 girls and 56 boys) between the ages of eight and thirteen had 126 restorations placed under strictly determined clinical conditions (LA, rubber dam, etching, lining etc). Unfortunately, two different composite resins were used and it is not clear from the study how much of a specific type was used. Fifty-seven of the original restorations were evaluated at three years i.e. 69 were lost to follow up. Only three restorations were noted to have failed during the time of the study. Unfortunately, there is little information that can be reliably gleaned from the study after three years, as nearly half of the original restorations had been lost to follow-up. The low re-examination rate, the clinical setting and the lack of information on the operators and examiners detract from the usefulness of this study.

Prospective study with historical controls and measuring “true” restoration failure (5, 5)

Van Meerberk et al, 1996.

This three year hospital / university based study carried out in Belgium was used to evaluate a number of dentine bonding systems in un-retentive Black's Class V

cavities which were subject to bevelling and etching or left alone. In total 420 Class V restorations were placed in 125 patients; neither tooth type nor baseline data were reported although the paper refers to classifying the cavity types by size. The authors do not provide the distribution of restorations in the dentition but do mention that the groups studied had similar numbers of restorations in them (100-110). The authors concluded from their study that systems that removed the smear layer and demineralised dentine appeared to be clinically superior to the other groups and that the newer systems showed improvements over the older ones despite the fact that the returns for each of the groups varied between 76 and 93% at three years; a difference of 17% on one occasion. The lack of group data, the setting and the use of rubber dam for Class V cavities reduce the generalisability of the findings to the practice environment. Additionally, it is noted that a team approach was used to evaluate the restorations, the number of people being involved not being stated in the paper.

Prospective study with concurrent controls using replacement criteria without training (6, 2)

Knibbs, 1992.

This two and a half year British study carried out at Newcastle University on dental students using one clinician and two evaluators (the operator plus one other) was designed to evaluate glass ionomer beneath Blacks Class I and II composite restorations. Thirty adult patients (17-23 years old; 14 male and 16 female) had 104 restorations placed in their permanent premolars or molars (68 were Class II and 36 Class I both with similar all the restorations were placed under similar conditions and used the same techniques. At the end of the study, only five

failures were reported. Unfortunately, the use of a single operator who was also an assessor reduces the usefulness of the findings as does the university setting. Approximately one third of the restorations were lost to follow up.

Prospective study with concurrent controls using replacement criteria with training (6, 3)

Letzel et al, 1987.

This two and half year Dutch study was designed to assess the influence that condensation instruments had on the clinical performance of amalgam restorations. Two clinicians and two evaluators were used in the study that was carried out in the general dental practice environment. The study involved the placement of 250 Blacks Class I or II restorations in the permanent teeth of 49 patients. Although rubber dam use was stated no other information with respect to placement technique was given. At the end of the study, all patients were still in the study with 238 restorations being evaluated. Twenty-eight failures were reported; one operator had twenty and the other eight. Although there was randomisation with respect to the packing techniques the study highlighted the differences that can exist between operators. The small number of operators and a lack of explanation on techniques or standards employed by the operators limits the generalisability of the study.

Prospective study with concurrent controls using validated outcome criteria (6, 4)

Hamilton et al, 1983.

This ten-year American study carried out by one clinician and two blinded evaluators in a clinical research facility compared the ability of two different amalgams to resist marginal failure. The study involved the assessment of 211 restorations (112 Spheraloy™ and 97 Dispersalloy™) placed in 77 patients over a period of ten years by one dentist. Although rubber dam use was stated there were little other controls placed on the use of the materials and the techniques utilised. The restorations were placed in permanent premolar and molar teeth but little is known about the types of cavities that were being restored. At one year, two years, three years, four years, five years and ten years the Dispersalloy™ amalgam showed less deterioration than the Spheraloy™. At ten years, there was no difference between the groups although there were only 44% of the restorations available for evaluation at the ten-year point. Although the study went to great lengths to separate marginal failure from degradation the results have to be considered as not being applicable to the environment of general dental practice due to the fact that only one clinician placed the restorations.

Prospective study with concurrent controls and measuring “true” restoration failure (6, 5)

Tyas, 1994.

This three-year Australian study carried out by one experienced operator evaluated the performance of Tenure™ in non-undercut Black's Class V cavities. Out of 40 original restorations, placed in five patients, 38 were available for

evaluation at the end of the trial. Only two of the restorations had failed. This study lacks generalisability due to the single operator who also acted as the evaluator. In addition, the materials are now not available for use. It is surprising how many clinical trials report on formulations and materials that for often-inexplicable reasons are removed and replaced by manufacturers for no apparent reason.

Clinical trial using replacement criteria without training (7,2)

Lidums et al, 1993.

This two-year Australian study carried out in a hospital by two clinicians and two evaluators was designed to compare a silver cermet, a posterior composite and a high copper amalgam in Black's Class I cavities in permanent teeth. Fifty-seven glass ionomers, 38 composites and 21 amalgams were placed in 7 premolars and 109 molars in 35 adult patients. The clinical conditions for placement were standardised. After two years, 4 of the 36 glass ionomers had failed. There were no failures in the amalgam or composite groups: seventeen and eight restorations remaining for analysis. The generalisability is affected by the clinical environment and the results obviously affected by the unexplained large drop out rate in the composite group.

Clinical trial using replacement criteria with training (7, 3)

Neo and Chew, 1996.

This three-year study carried out in a Singapore University by a single clinician and two trained and calibrated evaluators was designed to evaluate Black's class V restorations restored with glass ionomer, a composite resin and dentine bonding

agent or a composite resin / glass ionomer sandwich restoration. The study evaluated 159 restorations (100%) placed in eighteen adult patients. Fifty glass ionomers were placed in 19 incisors, 8 canines, 19 premolars and 5 molars. Fifty-five composites were placed in 19 incisors, 8 canines, 24 premolars and 4 molars. Fifty-four “sandwich” restorations were placed in 14 incisors, 12 canines, 23 premolars and 5 molars. All restorations were placed under standardised clinical conditions. Patients with a history of bruxism or clenching were excluded. After three years, there were two failures in the glass ionomer group, 12 failures in the composite group and 2 failures in the “sandwich” group. Statistical analysis showed a significantly greater failure rate in the composite group. This study lacks generalisability due to the setting in which it was carried out. There is also no way of confirming whether randomisation procedures were applied to the allocation of restorations to cavity.

Clinical trial using validated outcome criteria (7, 4)

Mjör and Jokstad, 1993.

This five-year public dental service Norwegian study carried out by three clinicians and three evaluators was designed to compare silver cermet, composite resin and amalgam in Black's Class II restorations in placed in the premolars (107) and molars (167) of 142 adolescents with a mean age of 13 years. Two hundred and seventy four restorations (88 amalgams, 95 cermets and 91 composites) were placed under standardised clinical conditions. After five years, 113 restorations (33 amalgams, 44 cermets and 36 composites) in 113 teeth (34 premolars, 79 molars) remained for analysis. The analysis of the results showed 4 amalgam failures, 22 cermet failures and 9 composite failures. This study fails to report

whether or not the evaluators were trained or calibrated. It states that secondary caries and bulk fracture were the commonest reasons for restoration failure but these findings cannot be corroborated or validated. There was a considerable variation in patient drop out rates between the operators; this influences the validity of the results. The study failed to mention whether or randomisation was applied for the treatments carried out.

Clinical trial measuring “true” restoration failure (7, 5)

Alhadainy and Abdalla, 1996.

This two year University based Egyptian study completed by two independent assessors was designed to evaluate the clinical performance of four adhesive systems in non-retentive Class V composite restorations. In this study loss of the restoration was the determinant of failure. Eighty class V restorations (four groups of twenty restorations) were placed under strict operative protocols (rubber dam, enamel bevelling, lining, etching, dentine bonding etc) in 38 patients of an unknown age and gender. After two years, 75 of the original restorations were still available for analysis that showed there to be no difference at all between the groups used in the study. This study was well controlled and showed an excellent follow up rate unfortunately the results cannot be applied to general dental practice it was carried out in a university. In addition, the relatively small numbers limit the statistical analysis.

Randomised controlled clinical trial using replacement criteria without training (8, 2)

Kilpatrick *et al*, 1995.

This two and one half year university based British study was designed to compare how a silver cermet cement and a glass ionomer cement performed in class II cavities in deciduous molars. The study was carried out by a single trained and calibrated examiner-operator. Thirty-seven children (21 male and 16 female) between the ages of four and ten years received 92 restorations (59 in first molars and 33 in second molars). The paper does not detail whether they were in maxillary or mandibular teeth. There was an equal split between the materials used. Unfortunately, little data is presented with respect to the standardisation of clinical technique. After eighteen months, the failure rate reported were 41% for the silver cermet and 23% for the glass ionomer. Mean survival data for the two materials are also presented (silver cermet 20.3 months and glass ionomer 25.3 months). The study concluded that silver cermet material should not be used in class II cavities in deciduous molars. Unfortunately this study used variable follow up times (4 to 31 months) and although presents valuable findings is not directly applicable to general practice because of the environment in which it was conducted and the use of a single operator-examiner.

Randomised controlled clinical trial using replacement criteria with training (8, 3)

Oldenburg *et al*, 1987.

This four year American based university study was designed to evaluate two composite resins and three cavity designs during the restoration of primary molar

teeth in young children. Three clinicians and three evaluators were used. Fifty children of unknown gender between the ages of four and eight were recruited into the study. Three hundred and fifty seven restorations (237 Class I, 188 Class II and 32 Class V) were placed with the cavity design and material type being designated randomly with the protocols for design etc being stated. After four years, 234 restorations were evaluated. No statistically significant results were recorded between the experimental groups but the authors reported that more of the minimal and the Class II restorations failed. Additionally they noted that the conventional preparation techniques seemed to fare better. As this study was carried out in a university setting its findings are not necessarily translatable to general dental practice.

Randomised controlled clinical trial using validated outcome criteria (8, 4)

Matis et al, 1996.

This ten-year American study carried out in a University setting was designed to compare two types of glass ionomer cement with composite resin in class V cavities. Thirty patients with an age range of 29-76 years took part in the study that involved premolar, canine and incisor teeth. The authors did not indicate whether the restorations were placed in the maxillary or mandibular arches. The clinical treatment protocols were well described and two trained and calibrated assessors evaluated the work of a single clinician. Unfortunately, this study exhibited a high drop-out rate at twelve months. At the end of the study, the authors showed that the glass ionomer cements showed statistically greater retention than the composite resin group. Again the findings are limited when it

comes to applying them to general dental practice because of the drop out rate, small numbers involved and setting.

Randomised controlled clinical trial measuring “true” restoration failure (8, 5)

Kilpatrick *et al*, 1996.

This twenty-seven month hospital based British trial was designed to assess the durability of a glass ionomer sealant restoration compared with a minimal composite restoration to treat occlusal caries in permanent teeth. Sixty-seven patients with paired class I cavities in their permanent premolars or molars were recruited into the study. The clinical parameters are quite well detailed in the paper and the numbers of pairs of restorations that had similar operative techniques applied to them are detailed. At the end of the study the author had 66 pairs of restorations available for analysis. Under the conditions of the trial no difference in the durability between the two materials was noted. Additionally, it appeared that the use of rubber dam had no effect on the results. However, as the trial was carried out by single operator and carried out in a hospital environment on a relatively small number of patients then the results are probably not applicable to general dental practice.

Exclusions

Equally important as to why certain papers were included in Chadwick *et al*'s (2001) systematic review are the reasons why some groups of papers were discarded or not included. The reasons for discarding some groups of studies are detailed below. However, there were also a number of types of study which do not

appear to have been carried out e.g. retrospective study with concurrent controls using replacement criteria with training (4,3), retrospective study with concurrent controls replacement using validated outcome criteria (4,4), retrospective study with concurrent controls and measuring “true” restoration failure (4,5), prospective study with historical controls using replacement criteria with training (5,3), prospective study with historical controls using validated outcome criteria (5,4).

There are also publications that cannot be “pigeon-holed” in the truest sense and which are excluded from systematic reviews but which present extremely valuable observations. An example of such is a series of papers published by Lucarotti *et al* (2005a, b and c) who present longitudinal data based on information collected by the Dental Estimates Board in of England and Wales (despite being published after Chadwick’s systematic review they highlight the point made). Lucarotti *et al* (2005c) showed in their analysis of over one half of a million restorations placed over a period of eleven years that the longest lasting restorations in England and Wales are most likely to be placed by young dentists in Wales who graduated from South Africa or Australasia! While this statement cannot be truly substantiated as it is derived from a statistical analysis, it does however command a certain credence as the data set used is essentially sound and powerful. These same researchers (Burke, 2006a; Lucarotti *et al*, 2005c) also state that gender may have little influence on restoration longevity but that age, charge-paying status and pattern of attendance in patients can play roles. Charge-paying status is also stated as a reason for increased restoration replacement rates in private practice by other researchers (Mjör, 1997; Marynuik, 1990; Marynuik and Kaplan, 1986) with replacement rates of up to 50% within ten years. The researchers above do, however, freely admit that more research is

needed to understand fully the reasons for such findings.

Justifications for not including papers with subjective opinion as the determinant of outcome

As mentioned earlier, subjective opinion limits the usefulness of a paper's findings as it is impossible to determine whether or not different clinicians looking at a restoration would reach the same conclusion.

Justifications for not including papers based on case reports in a systematic review

Case reports present the findings from one practitioner, working under a particular condition and using techniques that may be unique to them. It is extremely difficult to accept case studies as representing anything other than opinion as the effect modifiers involved in the study are often indeterminable and can represent significant deviation from what may be found in general dental practice.

2.3 Outlining the research: A statement of the perceived problem and overall purpose of the research

A significant part of work carried out in dental practice is replacement dentistry. As shown, research can be grouped through study design, outcome measure and by combining these (systematic review). However, research into restoration replacement is full of inconsistencies and it is difficult to carry out comparisons of what has been published due to differences in study design, outcome measure or reporting. For research to have value the study design and the outcome measure used to assess or measure a variable needs to be reliable and reproducible. The

reproducibility should be both between and within research groups. With specific reference to restoration replacement, the outcome measure should identify, ideally, an acceptable, agreeable and justifiable point of failure i.e. has a substantiated need for replacement. It is noted that a significant part of this replacement dentistry is perpetuated by subjective decision and the considerable variation in decision making between dentists causes concerns that are shared by the public, researchers, dentists themselves and third parties who may fund the treatment. It is not surprising that there is a call for consistency in decision making in today's environment where spiralling costs and consumerism take on ever-increasing roles that are more significant. It is not unreasonable to say that everyone wants value for money, that every one wants to be assured that dental treatment is not over prescribed, that it is necessary and that what has been provided is fit for purpose. This is particularly true when the validity of restoration replacement is regularly questioned and there is no conclusive evidence (from any source) that "doubtful" restorations could not survive for a significant time if they were left alone. It is arguable that if there is no reason to justify the replacement of a restoration then it should not be replaced. Suggestions that micro-leakage does not lead to recurrent caries (Mjör *et al*, 2005), that recurrent caries is an ill-defined parameter that cannot be differentiated easily from a stained cavosurface margin (Mjör and Toffenetti, 1992; Tyas, 1991) and that the "ditching" of cavosurface margins does not result in recurrent caries (Mjör and Qvist, 1997; Mjör, 1995) provide us with enough evidence to question the replacement needs of a significant number of restorations. It has also been suggested that marginal gaps need to be significantly large i.e. $> 400\mu\text{m}$ (Kidd *et al*, 1995) before restoration replacement should be considered justifiable. The location of the restoration

deficiency needs also to be taken into account with accessible defects probably being less susceptible to change through carious attack than more inaccessible deficiencies. In addition, we know that the replacement of restorations affects the long-term morbidity of the tooth that no restoration is perfect and if the restoration spiral of replacement can be reduced then perhaps this is a good thing. One starting point in helping to allay such concerns may be in the area of restoration replacement. Unfortunately, little research can relate long-term clinical failure to the condition of restorations when they present. As we know that restorations do not last forever perhaps reassurance is needed to ensure that they are replaced only when necessary. The consistency in decision making amongst practising dentists needs to be acceptable and ideally the consistency is not only comparable within a practitioner but also between practitioners. The USPHS criteria, which have developed over the years, would appear to be a good starting point. Not only are the criteria relatively easy to understand they appear to be clinically acceptable and cover what is normally evaluated during the clinical examination.

The purpose of this study was to determine the effect that restoration evaluation training had on the decision-making and restoration replacement rates amongst a group of practising dental practitioners. Secondary to this was a determination of usefulness in using the USPHS criteria as a clinical diagnostic tool.

Section 2.4 TABLES FOR CHAPTER 2

Table 2.1 Reasons stated in the literature for replacing restorations

	Rec	Ana	Pain	M Deg	M Disc	Wear	Loss	Tooth #	Rest #	Chnge	Disc	Endo	Unel	Pros
Braga et al, 2007	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Mjör et al, 2002	✓			✓	✓			✓	✓		✓			
Al Negrish, 2001	✓		✓	✓	✓		✓	✓	✓			✓		
Deligeorgi et al, 2000	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓			
Mjör et al, 2000b	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓			
Wilson et al, 1987	✓		✓	✓		✓		✓	✓				✓	
Friedl et al, 1994	✓	✓	✓	✓	✓		✓	✓	✓					
Pink et al, 1994	✓		✓	✓	✓	✓	✓	✓	✓		✓			
Mjör and Um 1993	✓	✓	✓	✓	✓			✓	✓		✓			
York and Arthur, 1993;	✓		✓	✓	✓			✓	✓					✓
Mjör and Toffenetti, 1992a	✓		✓	✓	✓			✓	✓					
Mjör and Toffenetti, 1992b	✓	✓	✓	✓	✓			✓	✓	✓	✓			
Qvist et al, 1990a	✓	✓		✓	✓		✓	✓	✓					
Qvist et al, 1990b	✓	✓		✓	✓		✓	✓	✓					
Mjör, 1989	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Klausner et al, 1987	✓			✓	✓			✓	✓					
Qvist et al, 1986a	✓	✓		✓	✓		✓	✓	✓					
Qvist et al, 1986b	✓	✓		✓	✓		✓	✓	✓					
Boyd and Richardson, 1985	✓	✓	✓		✓		✓	✓		✓				

Key to table

Rec = recurrent caries

Pain = pain and or sensitivity

M disc = marginal discolouration

Loss = loss or looseness of restoration

Rest # = restoration fracture

Disc = body discolouration of restoration

Unsi = unsightly

Ana = anatomical form

M deg = marginal degradation

Wear = tooth wear

Tooth # = tooth fracture

Chnge = change of material

Endo = replaced for endodontic reasons

Pros = replaced for prosthodontic reasons

Table 2.2 An indication of the similarities in restoration replacement rates between different countries over the decades

Continent	Location	Country	Author(s)	Type of Study	Period of study / sample	No. of Dentists in study	Rest'ns in study	Replacement
Africa		Jordan	Al Negrish, 2001	XS p and r	1 month	213	3,166	45%
		Saudia Arabia	Mahmood <i>et al</i> , 2004	XS p and r	2 weeks	unknown	326	30%
		South Africa	Gibb, 1966	XS	unknown	1	100	44%
Americas	Northern America	Canada	MacInnis <i>et al</i> , 1991	XS p and r	30 days	34	2,280	46%
			Boyd and Richardson, 1985	XS r	5 days	108	3,662 (s)	76%
			Richardson and Boyd 1973	XS p and r	5 days	50	1,518	76%
			York and Arthur, 1993	XS p and r	2 weeks	88	4,633	40%
			Klausner and Charbeneau, 1985	XS p and r	2 weeks	122	5,392	43%
		USA	Moore and Stewart, 1967	XS p and r	907 dental charts	unknown	8,493	45%
			Braga <i>et al</i> , 2007	XS p and r	4 weeks	37	592	60%
			Mjör and Um, 1993	XS p and r	2 weeks	9	1,175	39%
			Friedl <i>et al</i> , 1994	XS p and r	1 month	102	5,240	53%
			Deligeorgi <i>et al</i> , 2004	XS p and r	3 months	unknown	2,620	37%
Europe	South America	Brazil	Mjör and Um, 1993	XS p and r	2 weeks	9	1,175	39%
		Korea	Friedl <i>et al</i> , 1994	XS p and r	1 month	102	5,240	53%
		Greece	Deligeorgi <i>et al</i> , 2004	XS p and r	3 months	unknown	2,620	37%
	Mainland Europe	Italy	Mjör and Toffenetti 1992a	XS p and r	2 weeks	62	1,935	41%
			Mjör and Toffenetti 1992b	XS p and r	2 weeks	62	1,025	48%
			Qvist <i>et al</i> , 1990a	XS p and r	3 weeks	341	4,932	62%
	Scandanavia	Denmark	Qvist <i>et al</i> , 1990b	XS p and r	3 weeks	341	2,542	
			Qvist <i>et al</i> , 1986a	XS p and r	3 weeks	338	6,052	61%
			Qvist <i>et al</i> , 1990b	XS p and r	3 weeks	338	883	
		Finland	Palotie and Vehkalahti, 2002	XS p and r	205 sets of clinical notes	unknown	1,969	40%
	UK	Iceland	Mjör <i>et al</i> , 2002	XS p and r	Up to 100 rest ^{ns}	243	24,429	47%
		Sweden	Mjör 1981	XS p and r	2 weeks	85	5,487	74%
			Deligeorgi <i>et al</i> , 2004	XS p and r	3 months	unknown	2,620	52%
			Burke <i>et al</i> , 1999	XS p and r	Up to 100 rest ^{ns}	73	9,031	51%
			Frost, 2002	XS p and r	6 months	1	779	53%
			Wilson <i>et al</i> , 1997	XS p and r	6 weeks	22	2,379	60%
			Nuttall and Elderton, 1983	clinical decision making		15	1,094 (s)	54%

Key

XS = cross sectional

p = placement of restorations

r = replacement of restorations

(s) = surfaces

Table 2.3 Format of restoration failure (Kreulen *et al*, 1998)

	Reason for replacement	Description
True failures	Isthmus fracture	<i>Complete fracture of the restoration between box and step, whether or not parts of the restoration remain in situ</i>
	Recurrent caries	<i>Caries clinically or radiographically detected adjacent to the restoration</i>
	Enamel fracture	<i>Fracture, or fracture lines, in the enamel adjacent to the restoration (horizontal or vertical),so that enlargement of the restoration is necessary</i>
False failures	Non related caries	<i>Caries clinically or radiographically detected, not adjacent to the restoration (e.g. in the mesial surface of a tooth with a disto-occlusal restoration)</i>
	Orthodontics	<i>Tooth extraction indicated by orthodontic reasons</i>
	Aesthetics / Health	<i>Replacement of a restoration on request of the patient</i>

Table 2.4 Cross sectional studies reporting data on the longevity of amalgam restorations

Group 1 **Cross sectional studies**

Group 1a

Cross sectional studies based on subjective reporting

Mjör 2002; Al Negrish, 2001; Mjör and Moorehead 2000; Burke et al, 1999; Mjör, 1997; Smales and Hawthorne, 1996; Friedl et al, 1994; Jokstad et al, 1994; Mjör and Toffenetti, 1992; Smales et al, 1991; Allander et al, 1990; Drake et al, 1990; Qvist et al 1990a and b; Qvist et al, 1986a and b; Boyd and Richardson, 1985; Klausner and Charbeneau, 1985; Mjör 1981; Dahl and Eriksen, 1978; Richardson and Boyd, 1973

Group 1b

Cross sectional studies based on defined criteria for clinical failure (local or modified USPHS)

Cichon, 1999; Smales and Hawthorne, 1997; Mjör 1997; Martin and Bader, 1997; Hawthorne and Smales, 1997; Mahmood and Smales, 1994; Paterson, 1994; Mjör, 1992; Smales, 1991; Bjertness and Sønne, 1990; Mjör et al, 1990; Crabb, 1981; Allan, 1977; Lavelle, 1976; Robinson, 1971; Allan, 1969

Group 1c

Cross sectional studies with clearly defined criteria for clinical failure e.g. USPHS

Roulet, 1997

Table 2.5 Cohort studies looking at the longevity of amalgam restorations

Group 2 Longitudinal studies

Group 2a

Longitudinal studies with defined criteria for clinical failure (local or modified USPHS)

Van Nieuwenhuysen et al, 2003; Plasmans et al, 1998; Kreulen et al, 1998; Wilson et al, 1996; Akerboom et al, 1993; Barr-Agholme et al, 1991; Osborne et al, 1991; Smales et al, 1990; Letzel, 1989; Bentley and Drake 1986; Bates and Douglas, 1980; Lavelle, 1976

Group 2b

Longitudinal studies with defined criteria for clinical failure e.g. USPHS

Kiremitci and Bolay, 2003; Summitt et al, 2001; Setcos, 1999; Mair, 1998; Collins et al, 1998; Mair, 1985; Mjör, 1993; Welbury and Murray, 1990

Table 2.6 Papers accepted into the systematic review by Chadwick *et al*, (2001)

Increasing strength of outcome measure 

Outcome measure		Subjective opinion on restoration replacement	Criterion based decision making	Criterion and training based decision making (pseudo USPHS)	Criterion, training, calibrated decision making (USPHS)	Failure without intervention	
Study design							
		1	2	3	4	5	Total
Learned bodies	1	Not included	Not included	Not included	Not included	Not included	
Case studies	1	Not included	Not included	Not included	Not included	Not included	
Retrospective case series	2	Not included	Not included	Not included	Not included	Not included	
Prospective case Series	3	Not included	24 Included	6 Included	16 Included	8 Included	54
Retrospective study with concurrent controls	4	Not included	1 Included	None identified	None identified	None identified	1
Prospective study with historical controls	5	Not included	1 Included	None identified	None identified	1 Included	2
Prospective study with concurrent controls	6	Not included	18 Included	4 Included	6 Included	5 Included	33
Other controlled trial	7	Not included	41 Included	25 Included	17 Included	8 Included	91
Randomised controlled trial	8	Not included	5 Included	1 Included	5 Included	3 Included	14
			90	36	44	25	195

CHAPTER 3
MATERIALS AND METHOD

3.1 Overall aim

The aim of this study was to determine the effect that restoration evaluation training had on restoration assessment and replacement decisions amongst a group of practising dentists. The null-hypothesis for this research being that a simple training programme would have no effects on restoration replacement decision making by a group of dentists; this being assessed by a number of parameters. However, the project had a number of clearly defined phases that depended on identifiable targets within them;

1. a preparatory phase that included:

- the collection of suitably restored teeth and manufacture of suitable models for use during the simulated clinical phase
- the recruitment and selection of dentists to take part in the study
- the setting of the gold standard for the simulated clinical phase
- the design and manufacture of training material to allow a group of selected dentists to receive a restoration evaluation training programme

2. a simulated clinical phase that included:

- a pre-training phase involving the assessment of the restorations in the models by the untrained dentists
- the training of the dentists in restoration assessment
- a post-training phase involving the re-assessment of the restorations in the constructed models by the trained (test) dentists

3. a pre-clinical preparatory phase that included:

- the recruitment of patients with suitable restorations
- the setting of the gold standard for the clinical phase

4. a clinical phase that included:

- the assessment of the restorations by the trained (test) and untrained (control) dentists
- confirmation that the patients and restorations recruited into the study came to no harm.

5. a study evaluation by the participants.

A flow diagram summarising the project is shown in Figure 3.1 and a modified Gantt chart depicts the time-line for the project (Figure 3.2). The clinical phase required both clinical and ethical approval and this was obtained from the Cardiff and Vale NHS Trust and Bro Taf Health Authority (Appendices 3.1a and b) prior to commencing any of the other phases of the project; the principal author being well aware of the difficulties that can be experienced when seeking ethical approval.

The statistical methodology for the project is detailed at the end of this section as they impacted directly on the understanding of the work and evaluations undertaken.

3.2 Preparatory phase

3.2a Selection of teeth and construction of models

The aim of the work in this phase was the manufacture and construction of models for examination under clinical conditions in dental manikins by the dentists taking part in the study.

Adult human teeth that had been extracted and stored in a mixture of distilled water and thymol were hand searched⁸ and a pool of restored premolar

⁸ This pool of teeth being collected by the dental school from dental practices throughout Wales and before the human tissue act had come into force.

and molar teeth collected; at this stage there was no selection based on tooth or type of intra-coronal restoration. The collected teeth were kept until needed at room temperature in 20 ml screw topped specimen bottles⁹ containing a solution of nine parts distilled water and one part neutral buffered disinfectant¹⁰. Individual teeth were then inspected with dental loupes¹¹ under x2.5 magnification by the principle investigator (RM) and their suitability for use and inclusion in the study determined. Teeth with multi-surface restorations (more than two surfaces), gross caries, gross discolouration or severe extrinsic staining were discarded. The resultant pool of teeth was then stored until needed at room temperature in specimen bottles containing the solution of distilled water and neutral buffered disinfectant as described above.

From the pool of collected teeth premolar and molar teeth were selected and set as anatomically correct as possible in a pink coloured condensation cured silicone¹² material placed in acrylic resin¹³ replica jaws (Figure 3.3). The replica jaws being made from agar¹⁴ moulds of modified A3-OK Frasco¹⁵ models. In total, 88 teeth with 111 restorations were set in seven mandibular and six maxillary artificial jaws; each jaw being uniquely identified through the allocation a non sequential five digit number. When more than one restoration was present in a tooth the restorations to be evaluated were clearly defined. The type of teeth used and nature and distribution of restorations are detailed in Figure 3.4.

⁹ Greiner Bio-one, Brunel Way, Stroudwater Business Park, Stonehouse, Gloucestershire, GL10 3SX.

¹⁰ Youngs Hospec, Youngs Detergents, Lancare Ltd, Unit 2, 64 Liverpool Road, Great Sankey, Warrington, Cheshire WA5 1QX.

¹¹ Keiller, Swiss loupes ltd, Barnhurst, New Barn Road, Lingfield, Kent DA3 7JB

¹² Coltène Lab Putty, Coltène / Whaledent AG, Feldwiesenstrasse 20, 9450, Albstätten, Switzerland.

¹³ Oracryl, Bracon, High Street, Etchingham, East Sussex, TN19 7AL.

¹⁴ Croform, Davis Schottlander and Davis Ltd, Letchworth, Hertfordshire, England.

¹⁵ Frasco, Franz Sachs +Co. GmbH, Medical Technology, Plastic Technology, PO Box 1244, D-88061, Tettnang, Germany

When not being used, the teeth and replica jaws were kept at room temperature in sealed polythene jiffy bags; the specimens being covered in 10 cm x 10 cm surgical gauze¹⁶ soaked in a neutral buffered disinfectant and distilled water solution.

3.2b Determining and setting the gold standard for restoration replacement

The aim of this part of the study was to determine the condition of the restorations used in the simulated clinical phases of the project. To achieve this, the replica jaws (presented in a random order) were mounted in a phantom head¹⁷ (Figure 3.5) and placed on a standard dental operating chair¹⁸ (Figure 3.6). The replica jaws, with restored mounted teeth were then independently examined by two experienced clinicians (BC and RM) and the restorative status of the restorations determined. The evaluations that these examiners undertook were based on the USPHS criteria (Appendix 1a -e) which both the examiners were fully conversant with. The evaluations were recorded by a scribe onto a pre-designed evaluation recording sheet (Appendix 2.1); one examiner (RM) used x 2.5 magnification loupes, the other (BC) had optically corrected 20:20 vision. In addition to noting the USPHS categorisations a decision as to whether or not a restoration should be replaced was also made and recorded. These evaluations and all subsequent evaluations in the study were carried out under standard clinical conditions with a standard operating light¹⁹, size 4 front surface plain dental mirror²⁰, triple syringe and number nine probe²¹ being made available to the operator.

¹⁶ Rocialle Health, Dales Manor Business Park, Sawston, Cambridge CB2 4TJ.

¹⁷ KaVo Dental GmbH, Bismarking 39, 88400 Biberach, Germany

¹⁸ KaVo Dental GmbH, Bismarking 39, 88400 Biberach, Germany

¹⁹ KaVo Dental GmbH, Bismarking 39, 88400 Biberach, Germany

¹³ Dentsply Ash Instruments, Hamm Moor Leane, Addlestone, Weybridge, Surrey, KT15 2SE.

¹⁴ Dentsply Ash Instruments, Hamm Moor Leane, Addlestone, Weybridge, Surrey, KT15 2SE.

A week later, one quarter of the restorations were randomly chosen and re-presented to the examiners for evaluation under the same clinical conditions as before. Once more, the results of the examiner's deliberations and the tooth's restorative status were recorded. All data was entered onto an Excel 5.0²² spreadsheet before being imported into the SPSS²³ 12 statistical package.

After the independent evaluations of the restorations were completed an agreement by consensus was reached for those restorations that the experienced examiners disagreed over, these restorations being highlighted by a statistical comparison of the results using the "EXACT" function in the ExcelTM software package. The evaluations and subsequent re-evaluations of the two experienced examiners served a number of purposes; the agreed determination on the restorative condition of the teeth included in the study and hence the determination of the "gold" standard with respect to restoration replacement, the opportunity to evaluate the calibration between the "gold standard" evaluators (intra - and inter-examiner agreement) and a subjective determination on an examiner's ability to recall assessed restorations after one week.

3.2c The recruitment and selection of dentists to take part in the study

The aim of this part of the project was for a group of untrained volunteer dentists to be recruited into the study. A verbal approach was made by the principal investigator (RM) to twelve full-time and eleven part-time dentists employed within the Adult Dental Health Directorate of the University Dental Hospital of Cardiff. After expressions of interest were ascertained, written invitations (including a confirmatory reply slip) was forwarded to the volunteers (Appendix 3.2a). From

²² Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-7329

²³ SPSS Inc, 233 S. Walker Drive, 11th Floor, Chicago, IL 60606-6307.

this pool of twenty-three dentists, sixteen dental practitioners of differing clinical experience, gender and age were randomly selected by drawing a plastic Scrabble^{TM24} tile inscribed with a number from a bag (the numbers corresponding to a numbered list of the volunteers); these volunteers were then invited to take part in the project (Appendix 3.2b). None of the practitioners had ever participated in any form of restoration evaluation programme. Following the unfortunate death of one of the participants, a seventeenth dentist was chosen (as above) from the pool of volunteers. All of the “selected” practitioners agreed to take part in the research.

Sixteen practitioners were used in the study in order to facilitate data handling and provide a meaningful sample to test a new hypothesis through data gathering and statistical analysis. This number was chosen as it facilitated even and manageable splits in the research protocol. It also allowed for the reasonable evaluation of a new technique in a sample which was necessarily constrained by resource, manageability, practicality and time. The principal researcher had also been involved in other similar research protocols which had shown the workability and manageability of such a group size (McAndrew *et al*, 1994).

3.2d The design and manufacture of training material

In order to provide training material for the dentists taking part in the post-training phases of the study a number of clinical photographs of dental restorations were taken by the principle investigator (RM) at chair side using an Epson Photo PC 3100Z digital camera²⁵ or by the audio visual arts department in the University

²⁴ J.W. Spear & Sons, PLC of Enfield, Middlesex, England

²⁵ Epson (UK) Ltd., Maylands Avenue, Hemel Hempstead, Herts., HP2 7TJ

Dental Hospital of Wales on an Olympus E-330 digital camera²⁶. Photographs were only taken after appropriate consent had been received from patients (Appendix 3.3). These photographs formed the basis of a photographic collection used to produce a tailored training programme based on the USPHS criteria for suitable training in restoration evaluation. From the photographic material suitable photographic representations that could be used in the delivery of the training programme of the specific USPHS evaluation criteria were chosen by the gold standard examiners (RM and BC) and these photographs used to produce a training booklet for use within the training programme (Appendix 2.3 a to e).

3.3 The simulated-clinical phase

The aims of the simulated-clinical phase were twofold;

- the assessment of the selected restorations by the sixteen untrained dentists and a comparison of intra and inter-examiner agreement,
- the random selection and training of half the untrained dentists to make a test group and the evaluation of the effect that this training had on agreement through the measurement and comparison of their intra and inter-examiner agreement following their assessment of the restorations used in the previous section of the study.

3.3a The assessment of the restorations by the untrained dentists

The recruited dentists viewed the restorations mounted in a phantom head manikin under conditions identical to those used by the gold standard assessors in the pre-simulated clinical phase as described above and were asked to decide

²⁶ Olympus (UK) Ltd., 2-8 Honduras Street, London EC1Y 0TX.

whether they would or would not replace the viewed restorations. The use of magnification aids was left to the discretion of the recruit. For these evaluations, the dentists were told to assume that the patient was fit and well, that there was no dental pain or discomfort, and that the restorations were to be evaluated individually. The teeth and jaws were presented in a random order every time they were used and the decisions voiced by the dentists recorded by a scribe onto a proforma (Appendix 2.2). In addition to the yes and no comments on restoration replacement, a reason as to why a restoration was to be replaced was noted. In this part of the study, the dentists used their own evaluation criteria; remember none had been trained in the use of the United States Public Health Service criteria (USPHS). The results were entered into the Exel™ software package and exported onto a SPSS™ data sheet.

In addition to basic demographic detail (gender and years qualified), place of work, whether or not magnification loupes were used and the time taken to evaluate the restorations was recorded. If loupes were used their magnification strength was noted.

At least one week later (and on some occasions months later) approximately one quarter of the previously assessed restorations were randomly re-presented and re-examined under the same conditions as before and the results recorded. The times to make the evaluations were noted. In addition to this, the times between evaluations were calculated.

The data were again entered into the SPSS 12™ statistical programme for the following analyses; intra examiner kappa statistic, inter-examiner kappa statistic in comparison to the gold standard, Dice's Coincidence index for

agreements and disagreements, specificity, sensitivity, positive predictive value and negative predictive values being calculated.

3.3b The selection and training of the dentists

The aim of this phase was the selection of eight of the original sixteen dentists and to assess the effect that training had on their diagnostic ability and intra and inter-examiner agreement. This trained (test) group was selected by drawing suitably coded ScrabbleTM tiles from a bag in the same way to that used previously. However, before selection could take place it was necessary to manufacture training material that would be suitable for the delivery of a training programme to the trainees.

3.3b.i The training programme

A training programme was delivered approximately six months after the last set of assessments in the pre-training simulated clinical phase was completed. One person (RM) delivered the same training programme on four separate occasions to the eight randomly selected trainees. The number of participants on each training session ranged from one to three on each occasion (one with one, two with two and one with three). The training programme involved a systematic explanation of the USPHS criteria and the various assessment parameters in restoration evaluation; caries, restoration margin, margin discolouration, colour match and anatomical form. The presentation was reinforced by providing the trainee assessors with individual colour illustrated spiral bound and laminated booklets depicting and explaining the evaluation process to be used for the study (Appendix 2.3a-e).

After delivering this didactic component of the training a series of colour photographs (Appendix 2.4) were evaluated alongside the trainer before introducing the trainees to a model for examination. The participants were given assistance and guidance to ensure a full understanding of the evaluation criteria. In this instance the evaluation criteria were related to the trainers initials in order to mentally re-enforce them to the participants (**RMcA**; Recurrent cares, **M**arginal adaptation and **M**arginal, colour (aesthetics) and **A**natomical form. The training sessions lasted at least 45 minutes and no longer than one hour.

After training the trainees were given personal copies of the assessment plates for reference and asked to review the criteria before the next set of restoration assessments were made. It was re-enforced to the trainees that if they became uncertain of any of the assessment criteria that further assistance would be given and that models to practice on were available for their use.

3.3b.ii Evaluating the effects of training

Using the USPHS evaluation criteria, the test (trained) dentists completed an evaluation of 105 of the original 111 restorations (one model had been randomly removed and was used as a training model) at least one week after the training programme (as mentioned above this time frame gave the participants opportunity to reflect on the training received). After at least another week one quarter of the restorations were randomly re-presented and re-evaluated by the trained dentists. During these assessments colour plates of the assessment procedure were made available and used for revision purposes if needed. The evaluations were carried out under identical conditions as to that used in the pre-training simulated clinical

phase the trainees were reminded, when necessary, if they had used magnification loupes in the pre-training simulated clinical phase.

The results, as before, were entered by a scribe onto pre-designed proformas (Appendix 2.2) before being transcribed into a SPSS™ 12 data sheet for statistical analysis. The following was calculated; Dice's coincidence index for agreement and disagreement, the intra-examiner kappa statistic, the inter-examiner kappa statistic thorough comparison with the gold standard, sensitivity, specificity, positive predictive value and negative predictive value.

3.4 The pre-clinical phase: including the recruitment of patients and the determination of the gold standard for the clinical phase

The aim of the pre-clinical phase was the identification and recruitment of suitable patients with dental restorations of varying condition that could be evaluated by the sixteen original volunteer dentists; encompassing the recruits and the trainees.

A number of employees at the University Dental Hospital of Wales were approached by the principal researcher (RM) and asked if they would like to take part in a clinical trial which was looking at dental restorations and their replacement rates. Inclusion criteria included: having a number of plastic dental restorations, no removable prostheses, no dental pain or oro-facial discomfort and not actively undertaking any dental treatment by their general dental practitioner. Employees who felt they fulfilled these criteria were invited to attend a screening appointment. The screening appointment was carried out simultaneously by two experienced clinicians (BC and RM). A total of twenty patients were screened of which nine were found to be suitable for inclusion into the study. Once a volunteer was identified as potentially suitable for the study a further longer appointment was

made to determine and record the status of the patients' restorations by the same clinicians used to determine the gold standard of the laboratory phase restorations (BC and RM). As stated previously these examiners were fully conversant with the USPHS criteria which were again used to determine the gold standard for this part of the project. Only restorations in posterior teeth were evaluated. A total of 66 restorations were entered into the clinical phase of the study (Figure 3.7).

Once recruited and informed fully as to the nature of the study written consent was obtained from the patients (Appendix 3.3c -3.2d) and a number of dates set for the original sixteen volunteer dentists to examine the patients.

3.5 The clinical phase: including the assessment of the restorations recruited into the clinical phase by the trained and untrained dentists

The aim of the clinical study was to determine whether training of the volunteer dentists had produced differences between the two groups with respect to restoration assessments in the group of recruited patients.

The original sixteen volunteer dentists were contacted to determine their availabilities on a number of dates and suitable arrangements made for them to evaluate the patients under identical clinical conditions to all other phases in the research with a size 4 front surface plane mirror, triple syringe, number 9 dental probe and a standard operating dental light being made available. From the original sixteen assessors fourteen were able to participate in the clinical part of the study. This phase of the study took place approximately one month after the laboratory phase had been completed.

During the clinical examinations, the dentists were told, once again, to assume that the patient they were looking at was fit and well, that there was no

dental pain or discomfort, and that the restorations were to be evaluated individually. Individual scribes were provided and the assessors' findings entered onto a proforma identical to the ones used during earlier parts of the study (Appendix 2.2). The dentists assessed each restoration with the scribe relaying the order in which the restorations were to be assessed. When they had finished examining one patient, they moved on to the next. The dentists were given as much time as they felt necessary to complete their deliberations; the principal investigator noting the time taken to complete the examinations.

Two weeks later approximately one half of the restorations were re-examined. These examinations were carried out under the same conditions as all previous examinations.

As before, the results were entered into an ExelTM spreadsheet and then imported into the SPSSTM statistical analysis programme for analysis.

3.6 Post clinical phase

3.6a Confirmation of the gold standard and determination of no harm to the recruited patients

After the main study was completed, the patients and their restorations were re-examined by one of the principal investigators (RM) to confirm the integrity and fitness of the restorations examined during the clinical phase. Any restorations that required replacement were made known to the patients and suitable arrangements made.

3.6b An evaluation of procedure within the study by the participants

In order to evaluate this research from the perspective of the participants all assessors were asked to complete a questionnaire one month after the final examinations (Figure 3.8). Assessors were also given the option of a face-to-face meeting or structured telephone interview in order to note their answers to the questions posed in the questionnaire. The questionnaire was designed and developed by the principal author and “piloted” on two experienced colleagues before being used. The views of the non-trained assessors (control group) were also invited; these took place through the medium of informal, individual, discussions.

3.7 Statistical considerations

A number of statistical parameters were used and examined in this research (Table 3.1) they are discussed here as they have relevance in the interpretation of the results.

After discussing the project with a statistician, an analysis of q-q plots²⁷ and the subsequent evaluation of the results it was decided that non parametric analyses would be the most appropriate tests for determining statistical significances. The Wilcoxon signed rank test and the Mann-Whitney U test were used. The reasons for employing non-parametric analyses stemmed from the relatively small number of assessors used in the project, the data distribution and the fact that discrete parameters were being evaluated. It is acknowledged that these non parametric tests are less powerful than their equivalent parametric tests.

²⁷ Q-Q plots give an indication of a variable's distribution against the quantiles of a number of test distributions. The resultant probability plots aid in the selection of appropriate statistical analyses. If the values cluster around a straight line a direct association is likely and parametric analyses are used. If they do not and transformation does not result in a straight line distribution then non parametric statistical tests are deemed to be more suitable for statistical analysis.

An example of a q-q plot showing the non parametric distribution of data is shown in Figure 3.9. This research set the significance levels at the traditional 0.05 level, however it must be remembered that failure to achieve statistical significance may also be a reflection of insufficient power in the research and this has been taken into consideration when interpreting the results. The temptation to adjust the alpha level (i.e. to 0.15 or even 0.01) in order to compensate for the small group sizes and discreteness of the data as suggested by Stevens (1996) was resisted.

Table 3.1 details the statistical measures used to examine and interpret the results of the study; as described below each of the parameters having value in the interpretation of the results.

Sensitivity and specificity are two operating characteristics that essentially represent probability values and can be used to indicate how accurate a diagnostic procedure is. Sensitivity and specificity are particularly useful in describing the results of procedures in a dichotomous way, e.g. should a restoration be replaced or not? Sensitivity and specificity are often reported together as they give an indication of instances when the diagnosis may have been correct or indeed, when they may have been wrong. A test with a sensitivity of 100% detects all the cases of the “disease” or “parameter”. A test with low sensitivity represents one with many missed diagnoses. Similarly, a test with near 100% specificity indicates that it is excellent for determining cases that do not show the “disease” or “parameter” i.e. it correctly determines that there is no disease. Ideally, a diagnostic test should be highly sensitive and highly specific and this can be achieved for diseases that are truly dichotomous in nature. However, it is known that sensitivity and specificity values as indicators of diagnostic accuracy in dentistry vary significantly e.g. caries diagnosis based on clinical examination 0.13 and 0.94

respectively (Vendonschotsch *et al*, 1992), caries diagnosis based on radiographic examination 0.73 and 0.97 respectively (Mileman *et al*, 1985), gingival redness 0.27 and 0.67 respectively (Haffajee *et al*, 1983), plaque scoring 0.47 and 0.65 respectively (Haffajee *et al*, 1983), bleeding on probing 0.29 and 0.88 respectively (Lange, 1991). This apparent lack of sensitivity and specificity is due to the rather continuous nature of many dental diseases that exhibit degrees of disease presence.

Unfortunately, sensitivity and specificity do not tell “the whole story” as they really only depend on the characteristics of the procedure being evaluated. They do not take into consideration the prevalence of the disease or indeed thresholds which may be set for the diagnosis of the disease being looked for. There are instances where it is advisable to set a low diagnostic threshold for disease detection to ensure all cases of a life-threatening disease are picked up during a screening process e.g. cancer screening. This high sensitivity does however increase the likelihood of false positive results because of an associated low specificity. The reverse of this can be achieved by setting a high threshold for disease diagnosis which may be required for confirmation of disease in conditions that are not so life threatening; the diagnosis of the condition being deliberately highly specific (with few false positives) in order to prevent patients pursuing unnecessary or even invasive or irreversible treatment e.g. Temporomandibular Joint Dysfunction (TMD).

In order to combat the deficiencies associated with sensitivity and specificity calculations it is useful to calculate positive and negative predictive values as these tests do take into consideration the prevalence of the disease being measured in a particular sample. A test, which returns a high value with respect to

positive predictability, is more diagnostically accurate in determining whether the parameter being evaluated is actually present than one that returns a low value. The converse stands for the negative predictive value. Both these evaluations are calculated by considering the number of true positive and false positives of the measurements when they are compared with the gold standard. How sensitivity, specificity, positive predictive value and negative predictive value are calculated is illustrated in Table 3.3.

Dice's Coincidence Index (Dice, 1945) was also calculated from the observations of the assessors. This index provides a measure of probability that one examiner similarly diagnoses the findings reported by another examiner. In this study, Dices' Coincidence Index was used to calculate the probabilities of examiners similarly scoring that restorations were sound or needed replacement. The two formulas were used for calculating these scenarios to the gold standard determinations (Table 3.4).

Reliability in clinical examinations is measured by repeatability and reproducibility (Glazer, 1995) with a reliable examination being determined as one that gives consistent and dependable results with a minimal amount of error.

Despite its problems, Cohen's kappa statistic (Landis and Koch, 1977) is commonly used in the measurement, analysis and reporting of reliability of clinical evaluations and determinations. It is a measure of the degree of non-random agreement (i.e. non-chance agreement) between measurements of the variable. Table 3.2 indicates the related estimates of strength of agreement suggested by Landis and Koch (1977). These categories, although purely arbitrary, are well accepted as reasonable benchmarks for determining agreement amongst observers and their observations (Dunn and Everitt, 1995).

It is worth expanding on the perceived “problem” associated with Cohen’s kappa statistic. For kappa values to have real value the parameter being evaluated needs to have a reasonable spread in the sample, cohort, or population being examined. To explain, if only one person in a sample has a disease and the manifestation of this disease is obvious e.g. bright green skin, then it is clear that the condition being measured can be diagnosed easily, reliably and predictably and everyone participating in the trial would probably score a perfect kappa statistic. However, if the manifestation of the disease were less obvious e.g. microscopic petechiae, then determination of the disease without a microscope would be difficult and the subsequent evaluations would be poor. In this study, the main evaluation was whether or not a restoration needed replacing and this has been established as a difficult thing to do (Manhart, 2004). This would create difficulty in this research if the sample or population being examined had an unreasonably low number of failing restorations and the kappa statistic could suffer if examiners were able to clearly identify and remember the restorations that needed replacement. Equally, if the sample contained only restorations needing replacement then examiners would clearly remember this and, once again, return a near perfect kappa score.

Table 3.1 Statistical parameters calculated in the phases of the study

Determinant	Value and description in relation to the study
Sensitivity	A measure of the probability of correctly determining that a restoration should be replaced
Specificity	A measure of the probability of correctly determining that a restoration should not be replaced
Positive predictive value	The probability that a restoration does require replacement
Negative predictive value	The probability that a restoration does not require replacement
Dices' coincidence index	The probability that one examiners agreement on the condition of a restoration matches the assessment of another examiner
Kappa statistic for inter-examiner reliability	The degree of agreement amongst observers taking part in the study
Kappa statistic for intra-examiner reliability	The degree of agreement within observers taking part in the study

Table 3.2 Kappa values and related estimates of strength of agreement
(Landis and Koch, 1977)

Kappa value	Strength of agreement
0.00-0.10	Poor
0.11-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.8	Substantial
0.81-1.00	Almost perfect

Table 3.3 A 2x2 contingency table illustrating how positive and negative predictive values are calculated

		Gold Standard Result		
		Positive	Negative	Total
Procedure result	Positive	<i>True positive</i> (TP)	<i>False positive</i> (FP)	<i>TP + FP</i>
	Negative	<i>False negative</i> (FN)	<i>True negative</i> (TN)	<i>FN + TN</i>
	Total	<i>TP + FN</i>	<i>FP + TN</i>	<i>FN + TN + FP + TP</i>

$$\text{Sensitivity} = TP / (TP + FN)$$

$$\text{Specificity} = TN / (FP + TN)$$

$$\text{Positive predictive value} = TP / (TP + FP)$$

$$\text{Negative predictive value} = TN / (FN + TN)$$

Table 3.4 Explanation of how Dice's Coincidence Index is calculated

		Restorations		
		Examiners	Findings	
Gold standard determinations		Sound	Replace	Total
	Sound	<i>a</i>	<i>c</i>	<i>a + c</i>
	Replace	<i>b</i>	<i>d</i>	<i>b + d</i>
	Total	<i>a + b</i>	<i>c + d</i>	<i>a+b+c+d</i>

The formula for calculating the probability that a restoration diagnosed as sound by one examiner will be diagnosed similarly by another is

$$\frac{a}{[(a+b) + (a+c)]/2} = \frac{TP}{[(TP+FN) + (TP+FP)]/2}$$

The formula for calculating the probability that a restoration diagnosed as needing replacement by one examiner will be diagnosed similarly by another is

$$\frac{d}{[(c+d) + (b+d)]/2} = \frac{TN}{[(FP+TN) + (FN+TN)]/2}$$

TP = true positive

FP = false positive

TN = true negative

FN = false negative

3.9

FIGURES FOR CHAPTER 3

MATERIALS AND METHODS

Figure 3.1 Summary of materials and method

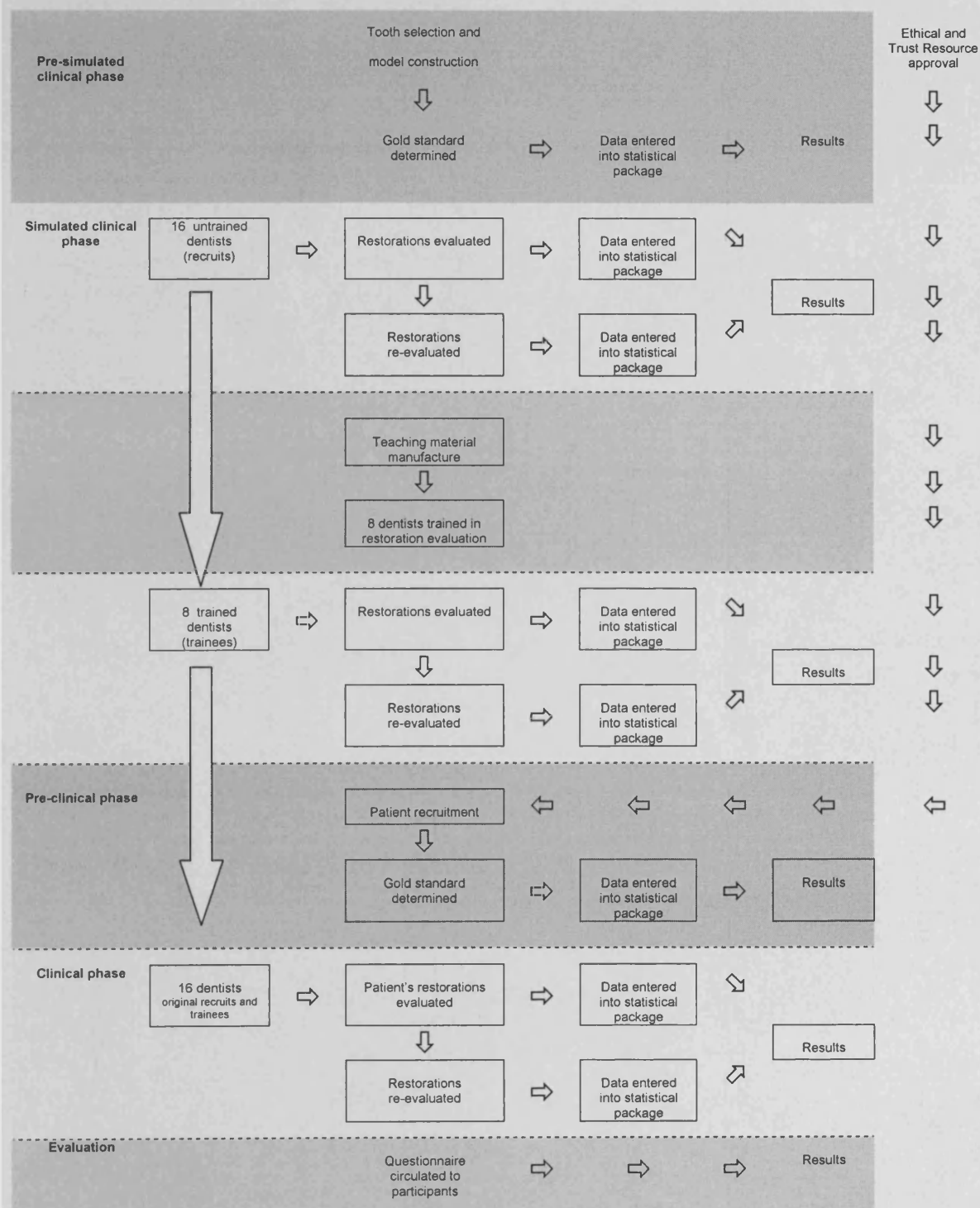


Figure 3.2 Timeline for study

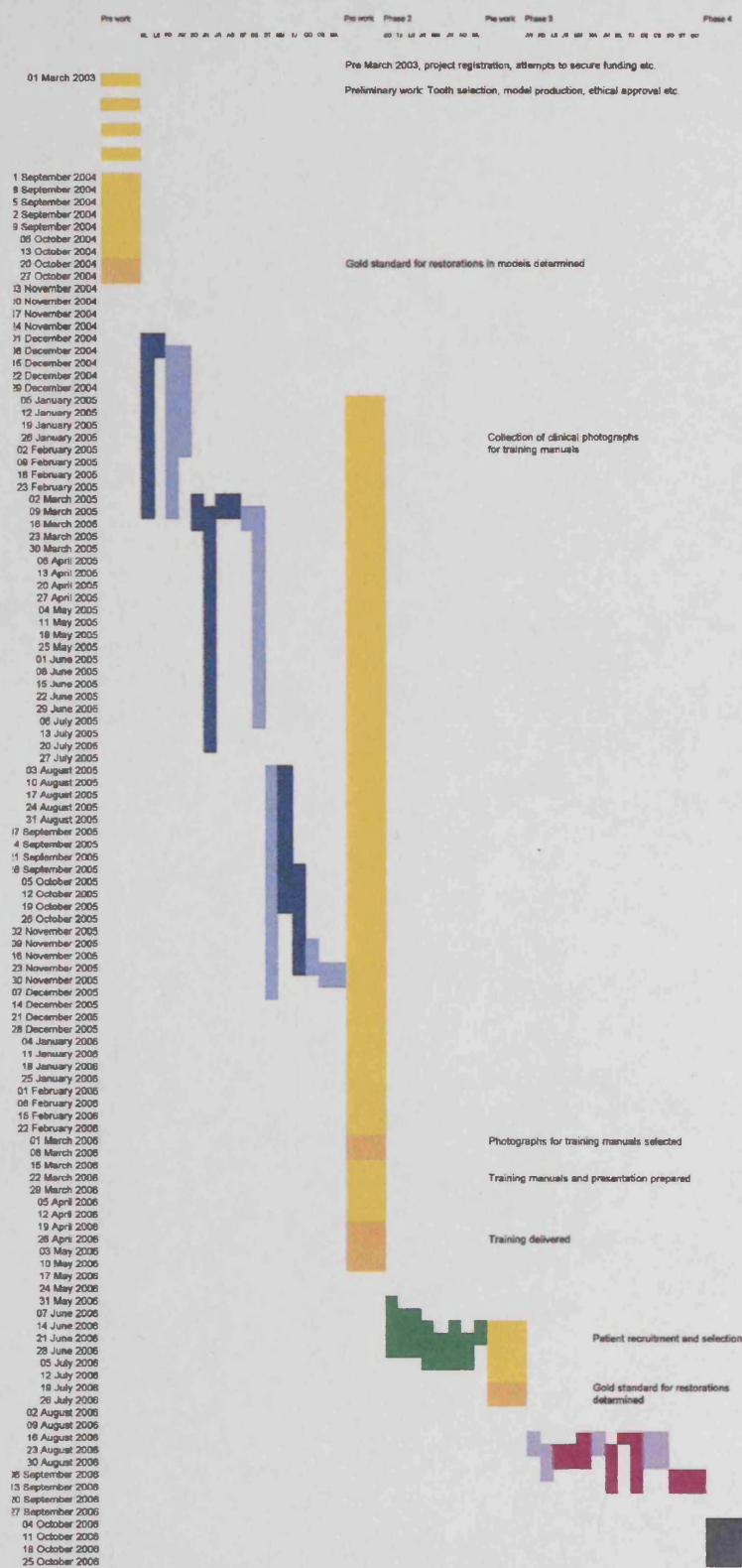


Figure 3.3 An example of replica jaws used in the study

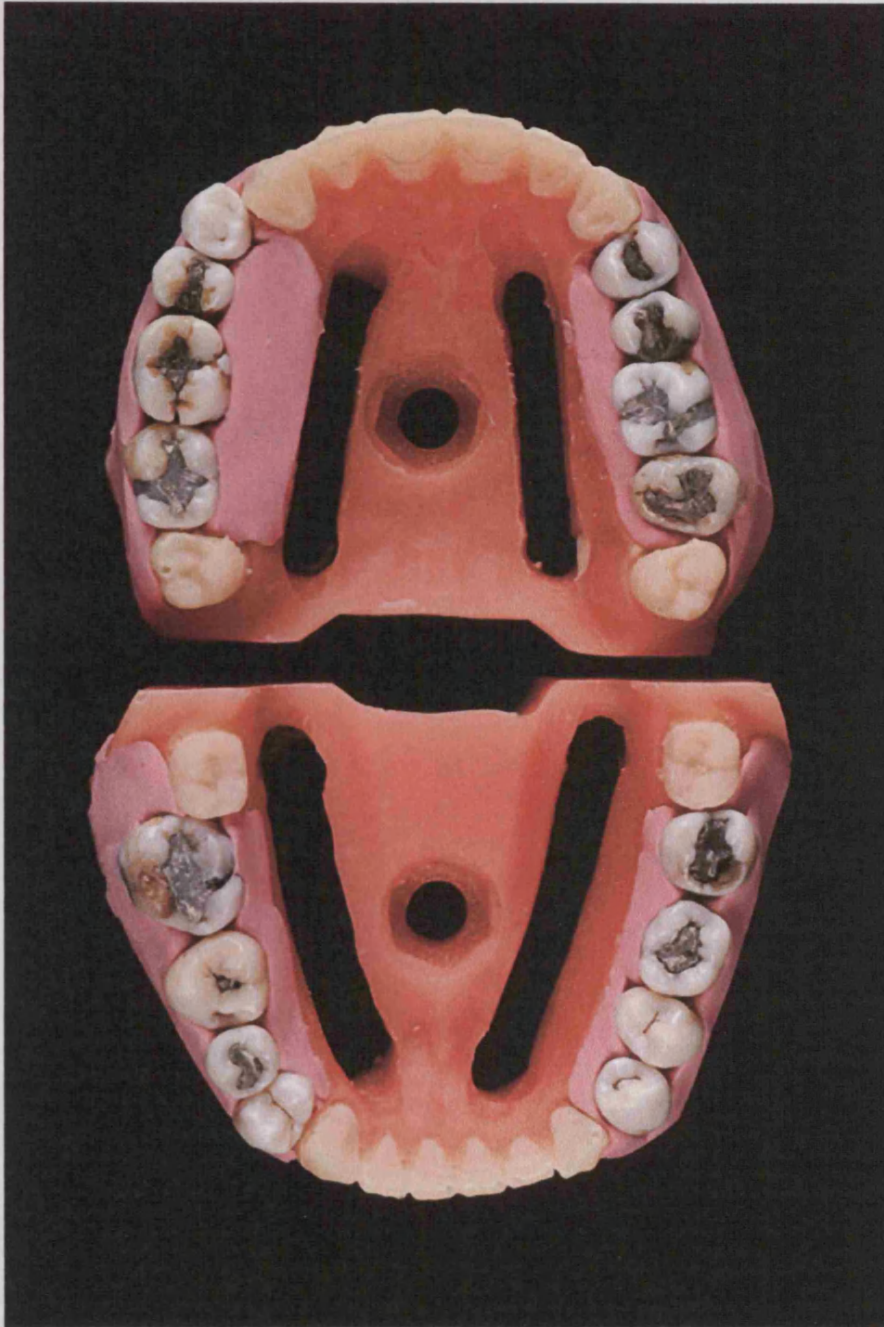


Figure 3.4 Type and distribution of restorations used for the laboratory phases

Tooth and restoration (s)								
Jaw (model number)	Right 7	Right 6	Right 5	Right 4	Left 4	Left 5	Left 6	Left 7
Maxilla 20793	O	O				O	O	
						B	OP	
Mandible 07019	O	O	O				O	O
	B							
Mandible 63428	O	O	DO	DO	DO	O	O	O
Maxilla 04537	O	O	DO	DO*	MO	MO	MO	OB
							B	
Mandible 67890	O	O	DO	DO	DO	MO	O	O
						DO		
Maxilla 57822	O	O	DO	MO	MO	DO	OB	O
								B
Mandible 14107	O	O	MO	DO	MO	M	OB	O
Maxilla 18104	O	O	DO		O	DO	O	O
	B						OP	
							B	
Mandible 58598	OL	OB	DO	MO	MO	MO	O	O
	B						M	
Mandible 59143	O	O	DO	DO	DO	DO	O	O
				B				
Maxilla 39156	O	O	O	MO	O	DO	OB	O
	DO	B						
Mandible 22564	OB*	O	DO	O		O	O	M*
				D				DO
								B
Mandible 25123	O	DO	MO*	MO	DO	DO	MOD	O

* tooth coloured

O occlusal

D distal

M mesial

L lingual

B buccal

Figure 3.5 Replica jaws in phantom head

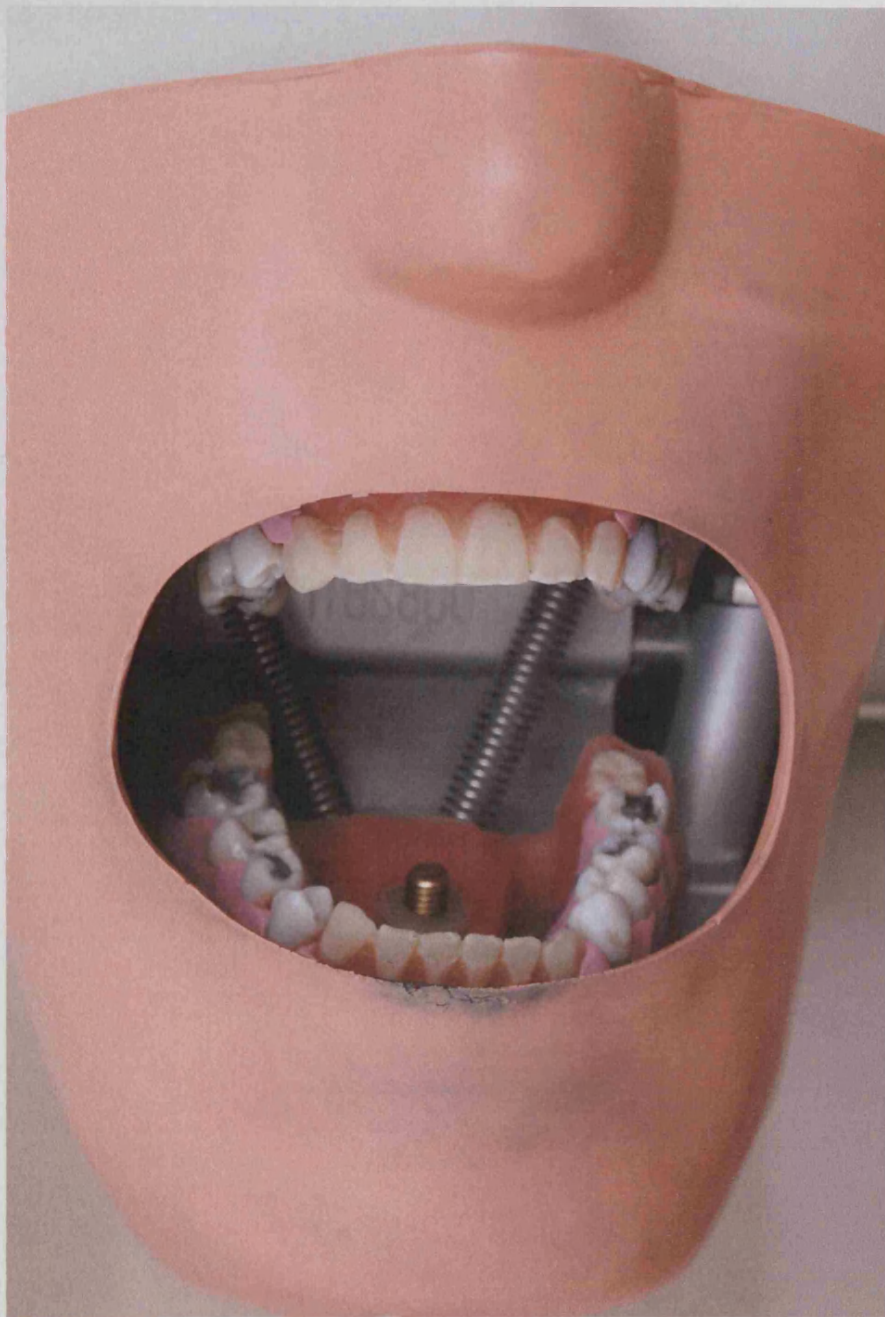


Figure 3.6 Phantom head with replica jaws and teeth being used in the clinic



Figure 3.7 Type and distribution of restorations in patients used for the clinical phase

		Tooth and restoration(s)								
Patient	Age		Right 7	Right 6	Right 5	Right 4	Left 4	Left 5	Left 6	Left 7
JH	35	Maxilla	DO ⁺ & O ⁺	O ⁺ & O ⁺					O ⁺ · O ⁺	O ⁺ · DO ⁺
		Mandible	O ⁺	O ⁺						DO ⁺
SS	30	Maxilla		MO	DO ⁺	DO ⁺			MO ⁺	MO ⁺
		Mandible		MO ⁺		DO ⁺		O ⁺	O ⁺	
AB	27	Maxilla	O	DO ⁺		DO ⁺	DO ⁺			
		Mandible							DO ⁺	
JC	48	Maxilla		MOB	MOD		DO	MOD ⁺	MOB	DO
		Mandible	MO	MODB	MOD ⁺			DO	MO	
RK	19	Maxilla	O	O					MOD	MO
		Mandible	MO	O					MOD	MODL
CM	23	Maxilla								
		Mandible	O ⁺	O ⁺	O ⁺				MO ⁺	
RJ	19	Maxilla	O						O ⁺	
		Mandible							OL	
MS	22	Maxilla						DO	MO ⁺ DO	MO
		Mandible	O	DO ⁺ &	O				DO	O
CK	54	Maxilla	DO						DO	MO
		Mandible	O	O					O	O

* tooth coloured

O occlusal

D distal

M mesial

L lingual

B buccal

Figure 3.8 Evaluation proforma

The clinical acceptability of the evaluation criteria

Section A: Place circle the response which you feel best reflects your answer

1. Overall, how would you rate the applicability of the evaluation criteria to clinical practice?	very easy easy difficult very difficult
---	--

2. How would you rate the applicability of the colour component of the evaluation criteria to clinical practice?	very easy easy difficult very difficult
--	--

3. How would you rate the applicability of the anatomical form component of the evaluation criteria to clinical practice?	very easy easy difficult very difficult
---	--

4. How would you rate the applicability of the marginal integrity component of the evaluation criteria to clinical practice?	very easy easy difficult very difficult
--	--

5. How would you rate the applicability of the caries component of the evaluation criteria to clinical practice?	very easy easy difficult very difficult
--	--

Section B: Please circle the response which best reflects your answer

1. Do you feel that the use of the evaluation criteria makes it easier for you to decide if a restoration needs replacing?	yes	no
--	-----	----

2. Do you feel your reliability and consistency with respect to restoration replacement need is improved through the application of the evaluation criteria?	yes	no
--	-----	----

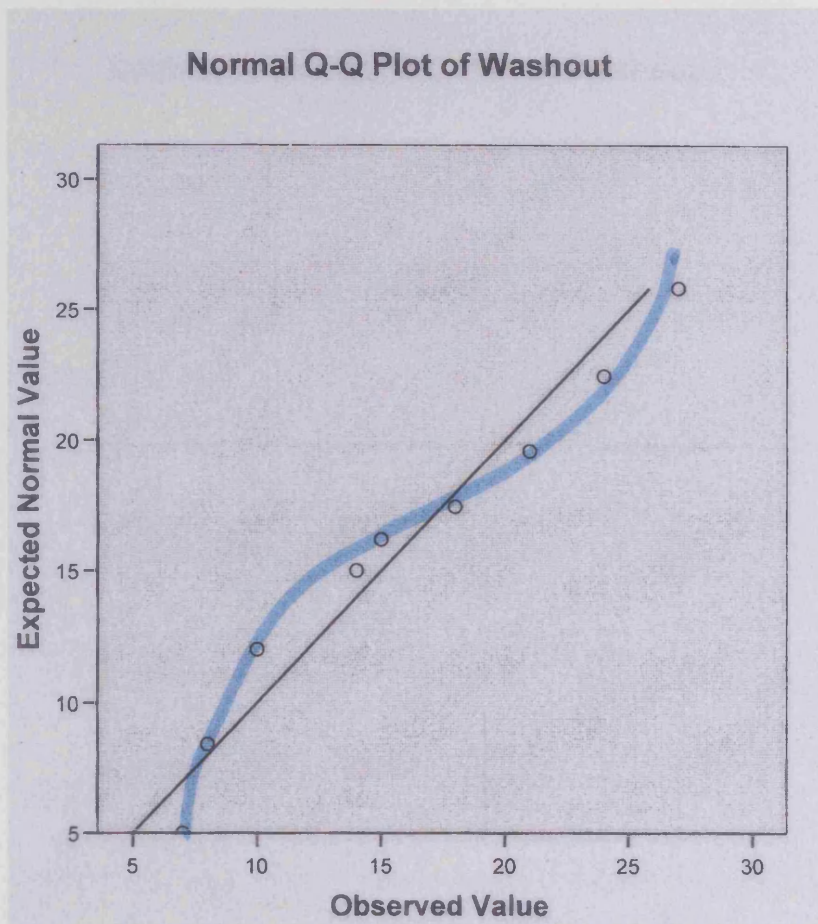
3. Do you think that these evaluation criteria have a role in clinical decision making?	yes	no
---	-----	----

4. Will you continue to use the evaluation criteria in your everyday working practice?	yes	no
--	-----	----

5. Has taking part in this project altered your clinical practice when it comes to decision making with respect to restoration replacement?	yes	no
---	-----	----

Please feel free to comment on this questionnaire or indeed any component of the project.
If you have any comments then feel free to write it on the reverse of this page.
Thank you for taking the time to complete this form. Please return it to me in the envelope provided.

Figure 3.9 Q-Q plot showing an example of the non-linear distribution of data. The graph shows the washout data but similar curves were noted for all parameters. An indication of how such distribution affects analysis is given as a footnote in Section 3.7



CHAPTER 4

RESULTS

4.1 Determining the gold standard for restoration replacement:

Assessments recorded and statistical analysis

To evaluate the decisions made by the dentists taking part in this research it was necessary to determine the restorative status of the restorations that were used in this study. In the simulated clinical and clinical phases, this was achieved from the assessments and deliberations of two experienced clinicians who were very familiar with the USPHS evaluation criteria (BC and RM, also referred to as the gold standard assessors). Their evaluations were made independently of each other under exactly the same conditions as all other examinations in the project. Tables 4.1.1a to 4.1.1d detail the gold standard assessors' deliberations on the restorations used in the simulated clinical phase; in order to measure reliability and calculate the degree of agreement all evaluations were duplicated.

The overall numbers of restorations scheduled for replacement in the simulated clinical phase by the gold standard assessors is presented in Table 4.1.2. The table shows that both assessors suggested that a similar number of restorations to be clinically unacceptable and require replacement (31 for BC and 32 for RM). The second set of evaluations scheduled fewer restorations for replacement; BC with 28 compared to 31 and RM with 27 compared to 32. It was noted that neither of the gold standard assessors scheduled a restoration for replacement on the second evaluation that had not been scheduled for replacement at the first evaluation by at least one of the assessors. Although the numbers of restorations scheduled for replacement by the gold assessors were similar the actual restorations were not the same. A comparison of the gold standard assessors' deliberations by combining the data collected from the first and second evaluations and utilising the EXACT function in ExcelTM identified that

areas where disagreement occurred and these findings are highlighted in Tables 4.1.1e and 4.4.1f; disagreements being shown by the return of the statement “FALSE” at a node. Out of the 555 decisions made by each of the gold standard assessors on the first assessment there were differences on 197 occasions. These differences were noted throughout the entire USPHS category groupings however, the differences were never greater than one character description in each of the category groupings (anatomy 39 of 111; caries 20 of 111; colour match 2 of 111; marginal integrity 54 of 111 and marginal discolouration 65 of 111). The disagreements highlighted seventeen restorations for replacement by one assessor and not the other. A similar process using the EXACT function in Excel™ revealed there to be 187 differences in opinion from the 555 evaluations made at the second evaluation and once again these differences were noted throughout the entire USPHS category groupings (anatomy 39 of 111; caries 20 of 111; colour match 2 of 111; marginal integrity 54 of 111 and marginal discolouration 64 of 111). However, this time the disagreements highlighted only eight restorations where the assessors differed. A consensus was reached on the restorative condition of the restorations highlighted by disagreements by the gold standard assessors carrying out a joint assessment of them (Figure 4.1 depicts some of these restorations). When this was completed it was decided that 28 of the 111 restorations needed replacement. There was no pattern between the examiners as to which restorations threw-up disagreements i.e. it could not be attributed to any particular examination criterion.

Tables 4.1.3 and 4.1.4 detail the results of the statistical analyses undertaken following the gold standard assessors evaluations; Table 4.1.3 details how they compared to themselves (intra-examiner variation), each other (inter-

examiner variation) and to the gold standard. There were no differences for the kappa values between the assessors when their first or second examinations were compared to the gold standard; 0.82(BC) and 0.77(RM) for the first examination and 0.95(BC) and 0.98(RM) for the second examination. These examinations revealed that the inter-examiner agreement with respect to the gold standard reached by consensus was in the higher range of the moderate and lower range of the substantial agreement range of Landis and Koch (1977) for the first examinations and reached the almost perfect range for the second evaluations. With respect to the replacement decisions made between the first and second examinations, the gold standard assessors showed very good agreement with themselves with the following intra examiner kappa values being returned; BC = 0.77, RM = 0.79.

No statistically significant differences were observed between the gold standard assessors for any of the parameters calculated (Tables 4.1.3 and 4.1.4). However, there was a noticeable difference between the positive and negative predictive values which suggested that the assessors found it easier to correctly predict a sound restoration than one that required replacement.

4.2 Basic data collected on the group participating in the study

Table 4.2.1 illustrates the demographic data collected for each of the dentists recruited into the project; an asterix indicates if they subsequently received training in restoration evaluation. It is important to note that, in order to reduce the introduction of unnecessary bias, in this study, participants were not allocated to a test or control (training or no further training) group until after the first simulated clinical phase of the project had been completed.

There was a relatively even split for the dentists in three of the four areas recorded; 8:8 for gender, 9:7 for magnification and 7:9 for experience. More hospital dentists were included in the study – ten compared to six mainly based in dental practices; this apparent anomaly resulted from the death of one of the study participants and that his replacement (a random allocation from the original group of volunteers) was a hospital based dentist; otherwise the split would have been 9:7. Overall, there appeared to be a good spread of experience (from 4 to 40 years with the mean years of experience being over 21 and one half years). A graph of the years since qualification confirms the study's volunteer dentists to be towards the experienced range (Figure 4.2)..

In this study, the volunteer dentists were neither discouraged nor encouraged to use visual aids (magnification) but merely advised to replicate what they would normally do in clinical practice. However, they were told that for the purposes of this study they would have to be consistent in their use of magnification and that if it was used then it would need to be used throughout. There was a range of magnification used by the study participants (x2 – x4.6). The distribution of magnification use was relatively evenly spread when looked at by gender (5 of 8 being male and using magnification with 4 of 8 being female) and years qualified (4 of 7 being qualified for >25 years and using magnification with 4 of 9 being qualified < 25 years). However, it appeared that a greater proportion of the hospital based dentists used magnification (7 out of 10) when compared to the dental practitioners (2 out of 6). It is uncertain whether or not this tendency towards hospital dentists being more familiar with the use of chair side magnification is represented in the practicing dental population as a whole.

The analyses of the demographic suggested that there was heterogeneity in the study's volunteer dentists.

4.3 Simulated clinical phases

4.3a Pre-training simulated clinical phase: assessments recorded and statistical analysis

The pre-training simulated clinical phase involved 16 of the original 22 volunteer dentists. No volunteer dentist involved in the study had received or undertaken any form of restoration evaluation training before the simulated clinical phase of the project. Each and every dentist evaluated all 111 restorations and after a washout period that ranged from 7 to 138 days they further evaluated a random sample of between 20 and 25% of the original 111 restorations used in the first simulated clinical phase; so that examination reliabilities could be determined with the kappa statistic. The raw data, a statement of ranges and the calculated mean with standard deviation is detailed in Table 4.3. Deliberations on restoration integrity ranged from a low of 25 to a high of 47 with the average being 35. The time to complete the first and second pre-training laboratory evaluations ranged from 25 to 31 minutes and eight to fifteen minutes respectively. The time period between the first simulated clinical evaluation and the repeat simulated clinical evaluation (so that the kappa statistic could be calculated) varied considerably; seven to 138 days.

After the data had been collected and entered into a computer database a number of statistical evaluations were undertaken; diagnostic accuracy was assessed through the calculation of sensitivity, specificity, positive predictive value

and negative predictive values and this data for individuals along with the mean values and standard deviations of the group are presented in Tables 4.3.1 - 4.3.3.

A further measure of diagnostic accuracy on whether a restoration would be evaluated as sound or requiring replacement was evaluated with the Dices' Coincidence Index. The mean Dices' coincidence index values for the group as a whole were closer to 1 which suggests greater diagnostic accuracy in the diagnosis of a sound restoration (0.88 ± 0.03); the scores returned for scheduling a restoration to be replaced were lower which suggests slightly greater difficulty in reaching this decision (0.70 ± 0.05). This finding is also represented with the higher specificity (0.92 ± 0.03) and lower sensitivity value (0.64 ± 0.09) and higher negative predictive value (0.84 ± 0.07) when compared with the positive predictive value (0.78 ± 0.09). All these findings indicated that the group were more likely to correctly diagnose a restoration to be sound than require replacement.

In addition to the above, the intra-examiner kappa statistic and the inter-examiner variation to the gold standard were calculated. The mean group results indicated that the study participants were more likely to agree with their own assessments on restoration viability than those determined by the gold standard assessors; the mean intra-examiner statistic (0.68 ± 0.09) being higher than that calculated for the mean kappa statistic for agreement with the gold standard (0.58 ± 0.07). The results indicated that only one of the volunteers returned an intra-examiner kappa statistic greater than that achieved by the gold standard assessors ($\pi @ 0.85$) and one matched that returned by BC ($\beta @ 0.77$).

Tables 4.3.1 to 4.3.3 (with associated figures) present a number of analyses completed on the demographic data represented in the volunteer dentists; gender, main place of work, use of magnification and years qualified.

Table 4.3.1 and Figure 4.3.1a detail the number of restorations scheduled for replacement during the pre-training simulated clinical examination. The analysis showed that there were no significant differences in number of restorations scheduled for replacement when examined by gender, use of magnification, place of work or whether or not a person had been qualified for more or less than 25 years. Similarly, no significant difference was noted in a number of the other parameters examined during the pre-training simulated clinical phase; time to complete the first examination (Table 4.3.1 and Figure 4.3.1b), time to complete the kappa statistic examination (Table 4.3.1 and Figure 4.3.1c), Dices' coincidence index calculations (Table 4.3.2 and Figure 4.3.2a), specificity (Table 4.3.2 and Figure 4.3.2b), positive predictive value (Table 4.3.3 and Figure 4.3.3.a), negative predictive value (Table 4.3.3 and Figure 4.3.3a), intra-examiner kappa statistic values (Table 4.3.3 and Figure 4.3.3b) and kappa statistic values when measure against the gold standard (Table 4.3.3 and Figure 4.3.3b). A more detailed observation of the findings failed to suggest that these areas necessitated further analysis and no trends were noticeable.

However, when analysing the washout period between the first and second pre-training simulated clinical examinations (Table 4.3.1 and Figure 4.3.1d) it was noted that there was a statistically significant difference in the time period between the first and second examinations when the hospital group was compared to the practitioner group (74.7 ± 45.6 days for the hospital group compared to 22.8 ± 18.3 days for the practitioner group, $p=0.02$).

There was also one other statistically significant result observed which related to the sensitivity calculations assessed by gender; $p=0.05$ (Table 4.3.2 and

Figure 4.3.2b). There being a suggestion that the female assessors were more likely to correctly identify the restorations requiring replacement.

Although these results highlighted potentially significant differences in some fields an overall analysis of the results in this section did not reveal any particular demographic parameter as a potential source of bias or to act as a confounder.

The results are highly suggestive that, as a whole, the group of volunteer assessors were fairly uniform and that little if any differences were apparent between any of the volunteers when the collected information was evaluated. The analyses indicated that as long as the allocations to the test and control groups were made in a suitably random manner that the test and control groups would be evenly matched.

4.3b Pre-training simulated clinical phase: assessments recorded and statistical analysis but separated after allocation to training or control groups

A statistical analysis of the pre-training simulated clinical phase data for the test and control groups (detailed later) with respect to the generic group parameters was conducted after all the simulated clinical phases were completed (Table 4.3.4). There were no statistically significant differences noted between the test and control groups during the pre-training simulated clinical phases. This analysis confirmed that the allocation to test and control groups was equitable, statistically sound and without discernable bias.

4.3c Post-training simulated clinical phase: results after training the test group

After completing the training programme, the trained dentists re-assessed 105 of the original 111 restorations and their findings noted; one block of the original restorations served as a training model. As with the pre-training simulated clinical phases, the time to complete the simulated clinical examinations was noted and analysed after all the evaluations. Table 4.4.1 summarises the calculations for the trained dentists, it also shows the values calculated for this group pre-training to allow comparison. This table formed to base for the following analyses.

Table 4.4.2 details the differences in washout times between the trained group and the untrained (control) group for the pre-training simulated clinical examinations. The period between the complete and partial evaluations (to allow kappa statistics to be calculated) for this part of the study showed a marked difference to those recorded in the pre-training simulated clinical phase. In the pre-training simulated clinical phase the mean value for this parameter was 56.8 days. It dropped to 18.1 days in the post-training simulated clinical phase. These mean time periods were compared statistically ($p=0.058$); while not reaching statistical significance at the 0.05 level it can be seen that the difference was noticeable. Graphic representations of the differences in washout periods are shown (Figures 4.4.2a-b).

A significant difference in the mean times taken to complete the full examinations in the pre-training and post-training simulated clinical examination phases were noted (Table 4.3.4). These differences were significant both within and between groups ($p=0.012$ and 0.001 respectively); it should also be remembered that fewer restorations were being evaluated for the full examination

in the post-training simulated clinical phase (105 as opposed to 111 restorations). These findings indicated that the trained examiners were taking significantly longer time over their deliberations in the post-training simulated clinical phase; approximately 20 minutes longer. The results are shown graphically in Figures 4.4.3a-b.

An increase in the mean examination time was also observed in the partial (kappa determination) examinations of the trained assessors (Table 4.4.4). Once again this finding was significant within the group and when compared to the untrained (control) group ($p= 0.012$ and 0.008). These results once again suggested that the trained assessors were taking longer over their deliberations; the findings are shown graphically in Figures 4.4.4a-b.

Bearing in mind that fewer restorations were being evaluated in the post-training simulated clinical phase, the results showed that the trained (test) group would replace fewer restorations than that observed in their pre-training simulated clinical examination phases; an average of 30 restorations (21 to 44) for the trained (test) group in the post-training simulated clinical phase as opposed to 34 (25 to 46) in the pre-training simulated clinical phase; the pre-training range being 25 to 46 and the post-training range 21 to 44. Proportionally, the replacement rates were calculated as 32.5 % for the untrained (control) group, 30% for the untrained (test) group and 28.5% for the trained (test) group. There were no statistically significant differences observed when the trained (test) group were compared to their pre-training evaluations or indeed with those of the un-trained (control) group. These findings and the statistical results are detailed in Table 4.4.5 and Figures 4.4.5a-b.

Dices' coincidence index values were calculated for decisions as to whether a restoration was sound or required replacing and these can be seen in Tables 4.4.6 and Table 4.4.7. There were no significant differences noted between the pre-training and post-training simulated clinical examinations; graphic representations of the findings are shown in Figures 4.4.6a-b and 4.4.7a-b.

When examined, the sensitivity calculations (Table 4.4.8) highlighted a noticeable difference between the un-trained (control) group and the trained (test) group ($p= 0.021$). This finding suggested that the trained (test) group would be more likely to agree with the gold standard in the evaluation of the restorations requiring replacement.

A comparison of the mean specificity calculations (Table 4.4.9) showed there to be no significant differences within the test group before or after training - or indeed between the trained (test) group and un-trained (control) group during the pre-training simulated clinical evaluations. This suggested that it was easier for all the volunteers to spot a good restoration than agree on one that should be replaced.

That training could significantly affect the test group's ability to correctly identify a restoration for replacement did not seem to be substantiated during this part of the project and this was confirmed when the positive and negative predictive values were examined (Tables 4.4.10 and 4.4.11) and there appeared to be no significant differences between the groups before or after training. These findings are illustrated in Figures 4.4.10a -b and 4.4.11a-b.

An examination of the kappa statistic for intra-examiner diagnostic in the trained (test) group's variability revealed that these dentists appeared to more consistent in the deliberations on restoration replacement (Table 4.4.12). This

was confirmed by a significant Wilcoxon test for the trained group which evaluated their findings for the pre-training and post-training simulated clinical evaluations ($p= 0.012$). The mean inter-examiner kappa evaluating the difference between the trained and untrained group also returned statistical significance ($p= 0.008$). These results suggested that the trained (test) examiners were more likely to agree with the other trained examiners decisions than with the un-trained examiners decisions; graphical representation of this can be seen (Figures 4.4.12a–b).

As well as showing a tendency to agree with their own findings, the trained examiners were also more likely to agree more consistently with the gold standards on restoration replacement set by the gold standard assessors (Table 4.4.13). There were significant differences after training between the trained (test) group and the un-trained (control) group ($p= 0.0015$) and within the trained group when they were compared with their own findings recorded for the simulated clinical examinations before and after training ($p= 0.0017$). These findings suggested that training improved the trained (test) group's diagnostic reliability and a tendency to agree with the gold standard.

4.4 Clinical examination phase: assessments recorded and statistical analysis

This part of the study involved the clinical assessment of a number of restorations in a group of volunteer patients by the trained (test) and untrained (control) dentists who had taken part in the previous simulated clinical examinations. The patients were drawn from a pool of screened patients and their details can be seen

in Table 4.5.1. There were nine patients and 64 restorations in this part of the study.

A number of parameters were recorded (the same as detailed in the simulated clinical phases) and statistical calculations undertaken for the assessors, their evaluations, and the time to complete their evaluations (Table 4.5.2).

The mean results for the trained (test) and untrained groups who took part in this part of the project are shown in Table 4.5.3; with graphic representations of the significant results detailed in Figures 4.5.3a-d.

There was a significant difference between the assessments made on the restorations in the clinical phase by the trained and the un-trained dentists ($p=0.034$) (Figure 4.5.3a) with the untrained dentists scheduling 9.71 ± 3.15 restorations for replacement and the trained dentists 6.00 ± 3.06 . It was also noted that the trained dentists took less time to make their decisions in the clinical phase and that this finding appeared highly significant; $p=0.003$ (Figure 4.5.3c). The trained dentists took, on average, 27.9 ± 3.44 minutes to complete their examinations with the un-trained dentists taking 36.7 ± 3.64 minutes. This apparent increase in speed of examination in the trained group was also observed during the examinations that took place to calculate the inter-examiner kappa statistic (Figure 4.5.3d). Again, the difference being statistically significant ($p=0.011$).

There was no difference in washout times between the two groups with the average time for the untrained (control) group being 15.4 ± 6.67 days and the trained (test) group 13.4 ± 7.01 days.

Table 4.5.3 details the mean results for all other statistical calculations.

There were no significant differences in the Dices' coincidence index calculations, sensitivity, specificity, positive predictive value, negative predictive value or inter-examiner statistic. There appeared to be a much greater disparity between the mean sensitivity and specificity scores and mean positive and negative predictive value scores during this part of the study when it was compared to the simulated clinical phase. However, once again it appeared that the assessors were more likely to diagnose a sound restoration correctly.

When compared the agreement of the trained (test) and un-trained (control) examiners to the gold standard determined by the independent (expert) assessors was statistically significant ($p= 0.002$) with the trained group showing a more consistent agreement with the gold standard (Figure 4.5.3b).

4.5 Post participation evaluation

This part of the study was undertaken to determine and evaluate what the volunteer dentists thought of the training programme and their thoughts on the usefulness of the USPHS criteria in clinical practice. The questionnaire was divided into two parts and had space for free comment. Out of the trained assessors seven of the eight completed and returned the questionnaire and one assessor opted for a face-to-face meeting. The results and free comments were tabulated (Tables 4.6.1 and 4.6.2). Out of the control group, six were willing to comment on how participation in the research project had affected their views on restoration evaluation and replacement decision making. Although the questionnaire did not lend itself to statistical analysis, the respondents made a number of interesting points.

Six of the 8 assessors found the applicability of the various assessment criteria to be easy or very easy, none finding the assessment process difficult.

All the trained assessors found the assessment procedure to be useful and felt it to have value in restoration assessment with 7 of the 8 trainees saying that they intended to continue to use the criteria in day-to-day practice. At the point of completing the questionnaire, 5 of the 8 assessors said that taking part in the project had altered their clinical practice: it had focused their attention when evaluating restorations. Four of the eight assessors thought that the USPHS system would be valuable in an undergraduate training programme.

With reference to the untrained control group, three specifically enquired as to when restoration evaluation training could be made available to them.

CHAPTER 4 RESULTS

TABLES AND FIGURES

1st assessment of restorations used for the simulated clinical phase (BC)

146

1st assessment of restorations used for the simulated clinical phase (RM)

147

Table 4.1.1c 2nd assessment of restorations used for the simulated clinical phase (BC)

	14 10 7								25 12 3								18 10 4												
	R4	R5	R6	R7	L4	L5	L6	L7	R4	R5	R6	R7	L4	L5	L6	L7	R5	R6	R7	R7	L4	L5	L6	L6	L6	L7			
Colour match	C	C	C	C	C	C	C	C	C	C	C	C	C	C	B	C	C	C	O	B	C	C	C	OP	B	O			
Marginal discolouration	A	B	A	A	A	A	A	B	A	A	A	A	B	A	B	A	B	B	A	C	C	B	B	A	C	B			
Anatomical form	A	A	A	A	A	B	C	B	A	C	B	A	C	A	B	B	B	A	A	A	A	C	A	A	A	A			
Marginal adaptation	A	B	B	A	A	B	C	B	A	C	B	A	C	A	B	B	B	A	A	A	A	C	A	A	A	A			
Caries assessment	A	B	A	A	A	A	B	B	A	A	A	A	B	A	A	A	A	B	A	A	A	B	A	A	A	A			
Replace	N	Y	N	N	N	N	Y	Y	N	Y	N	N	Y	N	N	N	N	Y	N	N	N	Y	N	N	N	N			
	58 59 8								59 14 3								39 15 6												
	R4	R5	R6	R7	R7	L4	L5	L6	L6	L7	R4	R4	R5	R6	R7	L4	L5	L6	L7	R4	R5	R6	R6	R7	R7	L4	L5	L6	L7
				O	B			O	M		B	MO									O	B	DO	O					
Colour match	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
Marginal discolouration	A	A	B	B	A	B	A	C	B	B	C	B	C	C	C	A	B	B	B	B	B	A	C	A	B	A	C	B	
Anatomical form	A	A	C	A	A	A	A	C	A	A	B	A	C	B	B	A	B	A	B	A	A	A	A	A	A	A	A	C	A
Marginal adaptation	A	B	C	A	A	A	A	C	B	B	B	B	C	B	B	A	B	B	B	C	B	B	C	A	A	A	A	C	B
Caries assessment	A	A	B	A	A	A	A	B	A	A	A	A	B	B	B	A	A	A	A	A	B	A	B	A	A	A	A	B	B
Replace	N	N	Y	N	N	N	N	Y	N	N	Y	N	Y	Y	N	N	N	N	N	Y	Y	N	Y	N	N	N	N	Y	N
	22 56 4								20 79 3								70 19												
	R4	R4	R6	R7	L5	L6	L7	L7	L7	R6	R7	R5	R5	R6	R6	L5	L6	L7	L7	R6	R7	R4	R5	R6	R7	L4	L5	L6	L7
	D	O				B	DO	M				O	B	OP	O			O	B										
Colour match	B	C	C	C	C	C	C	C	B	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
Marginal discolouration	B	B	B	C	B	A	A	B	B	B	B	B	C	A	C	B	A	B	A	A	B	A	B	B	B	C	A	B	B
Anatomical form	A	B	B	A	A	A	A	A	A	A	B	B	A	B	C	A	A	A	A	A	B	C	B	B	B	A	A	B	A
Marginal adaptation	B	A	A	C	A	B	A	B	A	A	B	A	C	A	D	A	B	A	B	B	C	C	B	B	B	A	A	B	B
Caries assessment	A	A	A	B	A	A	A	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Replace	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	Y	N	N	N	N	N	N	Y	Y	Y	N	N	N	N	N
	45 37								67 89 0								57 88 2												
	R4	R5	R6	R7	L4	L5	L6	L6	L7	R4	R5	R6	R7	L4	L5	L5	L6	L7	L7	R4	R5	R6	R7	L4	L5	L6	L7	L7	
						MO	B								DO	MO										O	B		
Colour match	B	C	B	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
Marginal discolouration	B	B	B	B	A	A	B	C	B	B	C	A	B	A	A	A	A	B	C	A	B	B	B	B	B	C	C		
Anatomical form	B	A	A	A	A	A	A	A	A	B	B	A	A	A	A	A	A	B	A	B	A	B	A	A	B	C	A	A	
Marginal adaptation	A	B	A	A	A	A	B	C	A	C	B	A	A	A	A	A	A	A	B	B	B	A	A	C	B	C	B	B	
Caries assessment	A	A	A	A	A	A	B	B	A	B	A	A	A	A	A	A	A	A	B	A	A	A	A	B	A	A	A	B	
Replace	N	N	N	N	N	N	Y	Y	N	Y	N	N	N	N	N	N	N	N	N	Y	N	N	N	Y	N	Y	N	Y	

Table 4.1.1d

2nd assessment of restorations used for the simulated clinical phase (RM)

	14 10 7								25 12 3								18 10 4															
	R4	R5	R6	R7	L4	L5	L6	L7	R4	R5	R6	R7	L4	L5	L6	L7	R5	R6	R7	R7	L4	L5	L6	L6	L6	L7						
																			O	B			OP	B	O	C						
Colour match	C	C	C	C	C	C	C	C	C	C	C	C	C	C	B	C	C	C	C	C	C	C	C	C	C	B						
Marginal discolouration	B	C	A	B	A	A	B	A	A	C	B	A	C	A	A	A	A	C	A	A	A	B	A	A	A	A						
Anatomical form	A	A	A	A	A	A	B	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A						
Marginal adaptation	A	B	A	A	A	A	A	B	A	A	B	A	D	A	A	A	A	A	A	A	A	B	A	A	A	A						
Caries assessment	A	B	A	A	A	A	B	A	A	B	A	A	B	A	A	A	A	B	A	A	A	B	A	A	A	A						
Replace	N	Y	N	N	N	N	Y	N	N	Y	N	N	Y	N	N	N	N	Y	N	N	N	Y	N	N	N	N						
	58 59 8								59 14 3								39 15 6															
	R4	R5	R6	R7	R7	L4	L5	L6	L6	L7	R4	R4	R5	R6	R7	L4	L5	L6	L7	R4	R5	R6	R6	R7	R7	L4	L5	L6	L7			
					O	B				O	M		B	MO								O	B	DO	O							
Colour match	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C				
Marginal discolouration	A	A	C	A	A	A	A	B	C	B	C	A	A	A	A	A	A	A	B	C	A	A	C	A	A	A	C	A	A			
Anatomical form	A	A	B	A	A	A	A	A	A	A	B	B	B	A	A	A	A	A	A	A	A	A	B	A	A	A	A	B	A			
Marginal adaptation	A	A	D	A	A	A	A	D	A	A	B	A	B	A	A	A	A	A	A	A	A	A	B	A	A	A	A	D	A			
Caries assessment	A	A	B	A	A	A	A	B	B	A	B	A	B	B	A	A	A	A	A	B	A	A	B	A	A	A	A	B	A			
Replace	N	N	Y	N	N	N	N	Y	N	N	Y	N	Y	Y	N	N	N	N	N	Y	Y	N	Y	N	N	N	N	Y	N			
	22 56 4								20 79 3								70 19								63 42 8							
	R4	R4	R6	R7	L5	L6	L7	L7	L7	R6	R7	R5	R5	R6	R6	L5	L6	L7	L7	R6	R7	R4	R5	R6	R7	L4	L5	L6	L7			
	D	O				B	DO	M				O	B	OP	O			O	B													
Colour match	C	C	C	C	C	C	C	C	B	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C				
Marginal discolouration	B	C	A	A	A	A	A	C	A	A	A	B	C	C	A	A	A	C	A	C	B	A	A	A	A	A	C	A	A			
Anatomical form	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
Marginal adaptation	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A	A	A	A	A	A	D	D	B	A	A	A	A	A	A			
Caries assessment	A	B	B	A	A	A	A	B	A	A	A	A	B	B	A	A	A	A	A	B	A	B	B	A	A	A	A	A	A			
Replace	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	Y	N	N	N	N	N	Y	Y	Y	N	N	N	N	N	N			
	45 37								67 89 0								57 82 2															
	R4	R5	R6	R7	L4	L5	L6	L6	L7	R4	R5	R6	R7	L4	L5	L5	L6	L7	R4	R5	R6	R7	L4	L5	L6	L7	L7					
						MO	B									DO	MO									O	B					
Colour match	B	C	B	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C				
Marginal discolouration	A	C	A	A	C	A	B	B	B	C	C	B	B	A	A	A	A	A	C	A	A	A	B	A	B	B	C					
Anatomical form	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A					
Marginal adaptation	A	A	A	A	A	A	B	A	A	D	A	A	A	A	A	A	A	A	A	A	A	A	A	A	D	A	B					
Caries assessment	A	A	A	A	B	A	B	A	A	B	A	A	A	A	A	A	A	A	B	A	A	A	A	A	B	A	B					
Replace	N	N	N	N	N	N	Y	Y	N	Y	N	N	N	N	N	N	N	N	N	Y	N	N	N	Y	N	Y	N	Y				

Agreement and disagreements between the 1st assessments of the restorations made by the gold standard assessors (BC and RM). The “disagreements” are shown at the false nodes.

150

Agreement and disagreements between the 2nd assessments of the restorations made by the gold standard assessors (BC and RM). The “disagreements” are shown at the false nodes.

151

Table 4.1.2 Summary data for restoration replacements decisions made by the gold standard assessors on the restorations used in the simulated clinical phases of the study

Gold standard assessor	Number of teeth examined	Replacements scheduled (1 st examination)	Replacements scheduled (2 nd examination)
BC	111	31	28
RM	111	32	27

Table 4.1.3 Results achieved by the gold standard assessors when compared to themselves, each other and the agreed gold standard

Gold standard assessor	Dices coincidence Index for teeth requiring restoration replacement	Dices coincidence Index for sound restorations	Inter – examiner kappa statistic i.e. compared to the gold standard	Intra-examiner kappa statistic i.e. agreement between the first and second examinations
BC	0.89 ⁺ 0.95 [*]	0.96 ⁺ 0.98 [*]	0.82 ⁺ 0.95 [*]	0.77
RM	0.82 ⁺ 0.96 [*]	0.93 ⁺ 0.99 [*]	0.77 ⁺ 0.98 [*]	0.79

+ = first examination of restorations

* = second examination of restorations

Table 4.1.4 Results achieved by the gold standard assessors when compared to themselves and the agreed gold standard for restoration replacement

Gold standard assessor	Sensitivity	Specificity	Positive predictive value	Negative predictive value
BC compared to self	0.92	0.91	0.75	0.96
BC compared to gold standard	0.92 ⁺ 0.93 [*]	0.93 ⁺ 1.00 [*]	0.81 ⁺ 1.00 [*]	0.97 ⁺ 0.98 [*]
RM compared to self	0.93	0.92	0.78	0.97
RM compared to gold standard	0.89 ⁺ 0.96 [*]	0.92 ⁺ 1.00 [*]	0.78 ⁺ 1.00 [*]	0.96 ⁺ 0.98 [*]

⁺ first examination of restorations

^{*} second examination of restorations

Table 4.2.1

Basic demographic details for the randomly selected volunteers

Assessor	Gender	Years qualified	Main place of work	Magnification
A	M	25	Hospital	Spectacles
B	M	32	Hospital	4.6
γ^*	F	11	Hospital	2.5
δ^*	M	9	Hospital	2.4
ϵ^*	F	27	Hospital	Spectacles
ζ^*	M	21	Hospital	3
η^*	M	14	Hospital	None
θ^*	M	26	Hospital	3
I	F	5	Practice	2.5
K	M	40	Hospital	2.5
Λ	F	26	Hospital	2
μ^*	F	23	Practice	Spectacles
ν^*	F	29	Practice	2.6
Ξ	F	23	Practice	Spectacles
O	M	17	Practice	None
Π	F	22	Practice	None

* indicates subsequently selected for training

Generic category	Range	Mean $\pm$ Standard deviation	
Years qualified	4 – 40 years	21.63 $\pm$ 8.98	
Magnification	2 - 4.6	2.78 $\pm$ 7.44	
Ratios for assessors	Overall	Test	Control
Male : Female	8:8	4:4	4:4
Hospital : Practice	10:6	6:2	4:4
Magnification :no magnification	9:7	5:3	4:4
Qualified > 25 years: Qualified < 25 years	7:9	5:3	4:4

Table 4.3 Raw data resulting from the initial examination of the pooled restorations prior to any training during the simulated clinical phase (* indicates that an assessor subsequently received training)

Assessor	Washout (days)	Time taken to examine all restorations (minutes)	Time taken to complete examination for intra-examiner kappa statistic (minutes)	Number of restorations scheduled for replacement	Dices' Coincidence Statistic for agreement that a restoration was sound	Dices' Coincidence Statistic for agreement that a restoration required replacement	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Kappa self	Kappa gold
A	52	45	12	28	0.93	0.76	0.73	0.89	0.68	0.92	0.65	0.61
B	92	43	9	44	0.84	0.67	0.54	0.94	0.86	0.76	0.77	0.52
γ*	96	34	10	28	0.93	0.80	0.78	0.93	0.79	0.93	0.73	0.71
δ*	94	45	9	41	0.85	0.66	0.56	0.93	0.82	0.78	0.65	0.52
ε*	21	25	12	39	0.86	0.69	0.59	0.93	0.82	0.81	0.67	0.56
ζ*	15	42	11	38	0.86	0.67	0.58	0.92	0.79	0.81	0.69	0.53
η*	138	40	9	25	0.92	0.73	0.76	0.89	0.68	0.93	0.76	0.63
θ*	14	51	14	46	0.86	0.71	0.56	0.97	0.93	0.76	0.65	0.57
ι	14	35	8	29	0.92	0.77	0.76	0.93	0.79	0.92	0.52	0.69
κ	117	38	9	43	0.82	0.62	0.51	0.91	0.79	0.75	0.72	0.45
λ	108	38	14	34	0.89	0.71	0.65	0.92	0.79	0.86	0.75	0.59
μ*	21	45	12	29	0.88	0.67	0.66	0.89	0.68	0.88	0.69	0.55
ν*	56	31	8	26	0.88	0.63	0.65	0.87	0.61	0.89	0.51	0.51
ξ	7	40	11	30	0.90	0.72	0.70	0.91	0.75	0.89	0.68	0.63
ο	7	32	10	47	0.86	0.72	0.57	0.98	0.96	0.76	0.62	0.59
π	18	40	15	37	0.88	0.71	0.62	0.93	0.82	0.83	0.85	0.59

	Range	Mean $\pm$ SD
Washout period	7 – 138 days	54.4 $\pm$ 45.7
Time main exam	25 – 31 minutes	39 $\pm$ 6.49
Time kappa exam	8 – 15 minutes	10.8 $\pm$ 2.20
Restorations to replace [§]	25 - 47	35.3 $\pm$ 7.53
Agree sound	0.86 – 0.93	0.88 $\pm$ 0.03
Agree replace	0.62 – 0.80	0.70 $\pm$ 0.05
Sensitivity	0.51 – 0.78	0.64 $\pm$ 0.09
Specificity	0.87 – 0.98	0.92 $\pm$ 0.03
Positive predictive value	0.61 – 0.96	0.78 $\pm$ 0.09
Negative predictive value	0.75 – 0.93	0.84 $\pm$ 0.07
Intra-examiner kappa	0.52 – 0.85	0.68 $\pm$ 0.09
Kappa to gold standard	0.45 – 0.71	0.58 $\pm$ 0.07

[§] The gold standard operators agreed that there were 83 sound restorations and 28 restorations that would require replacement.

Table 4.3.1 Pre-training simulated clinical phase results evaluated by demographic parameters and expressed as mean, standard deviation and significance (Mann Whitney U test)

Demographic parameters	N	Number of restorations scheduled for replacement			p	Time taken to examine all restorations (minutes)			p	Time taken to complete Examination for intra-examiner kappa statistic (minutes)			p	Time between initial and re-examinations (days)		
		Mean	SD			Mean	SD			Mean	SD			Mean	SD	
Gender	Male	8	36.6	6.72	0.10	38.8	6.76	0.06		10.8	1.83	0.49		74.4	39.6	0.6
	Female	8	33.6	8.72		39.3	6.67			10.9	2.64			28.2	44.7	
Magnification	Yes	9	33.3	8.01	0.16	38.1	7.24	0.46		11.6	1.90	0.46		37.7	46.7	0.34
	No	7	36.6	7.55		39.7	6.20			10.2	2.33			39.1	43.0	
Workplace	Hospital	10	36.4	7.73	0.55	40.1	7.09	0.33		10.9	2.02	0.74		74.7	45.6	0.02*
	Practice	6	33.0	7.77		37.2	5.42			10.7	2.66			22.8	18.3	
Years qualified	<25 years	8	33.8	7.29	0.66	39.2	4.66	0.51		10.6	2.07	0.96		45.6	49.6	0.21
	>25 years	8	36.9	8.38		38.7	8.73			11.1	2.48			38.5	40.9	

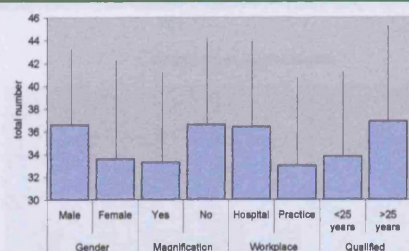


Figure 4.1a

Mean number of restorations scheduled for replacement during the pre-training simulated clinical phase

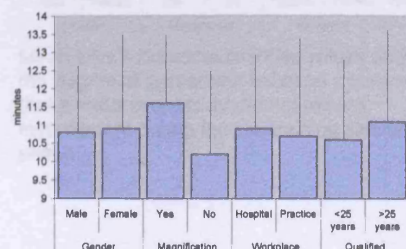


Figure 4.1c

Time taken to evaluate the restorations for the "kappa" determinations during the pre-training simulated clinical phase

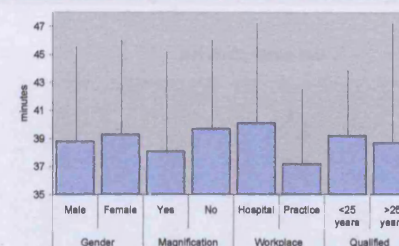


Figure 4.1b

Time taken to complete the initial examination phase during the pre-training simulated clinical phase

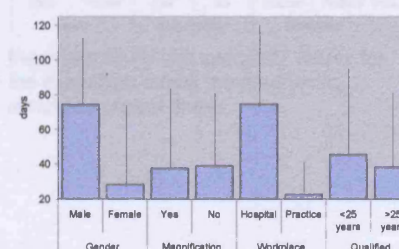


Figure 4.1d

Time between the initial and re-examinations during the pre-training simulated clinical phase

Table 4.3.2 Pre-training simulated clinical phase results evaluated by demographic parameters and expressed as mean, standard deviation and significance (Mann Whitney U test)

Demographic parameters	N	Dices' Coincidence Statistic for agreement that a restoration was sound			p	Dices' Coincidence Statistic for agreement that a restoration required replacement			Sensitivity			Specificity		
		Mean	SD			Mean	SD		Mean	SD		Mean	SD	
Gender	Male	8	0.87	0.04	0.44	0.69	0.05	0.1	0.6	0.09	0.05	0.93	0.03	0.46
	Female	8	0.89	0.03		0.71	0.05		0.68	0.09		0.91	0.02	
Magnification	Yes	9	0.87	0.04	0.24	0.69	0.06	0.16	0.62	0.1	0.17	0.92	0.03	0.17
	No	7	0.89	0.03		0.71	0.03		0.66	0.07		0.92	0.03	
Workplace	Hospital	10	0.88	0.04	0.44	0.7	0.05	0.55	0.63	0.1	0.33	0.92	0.02	0.66
	Practice	6	0.89	0.02		0.7	0.05		0.66	0.06		0.92	0.04	
Years qualified	<25 years	8	0.89	0.03	0.08	0.72	0.05	0.66	0.67	0.09	0.19	0.92	0.03	0.59
	>25 years	8	0.87	0.04		0.68	0.05		0.61	0.08		0.92	0.03	

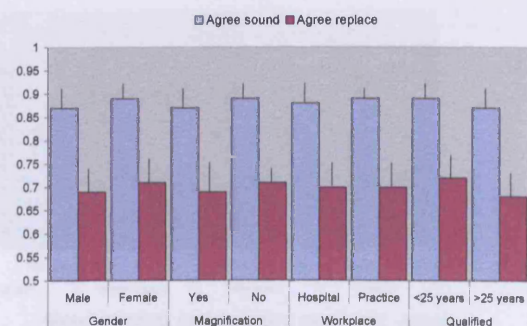


Figure 4.2 a

Mean Dice's coincidence index values showing the degree of agreement between examiners that a restoration did or did not require replacement during the pre-training simulated clinical phase

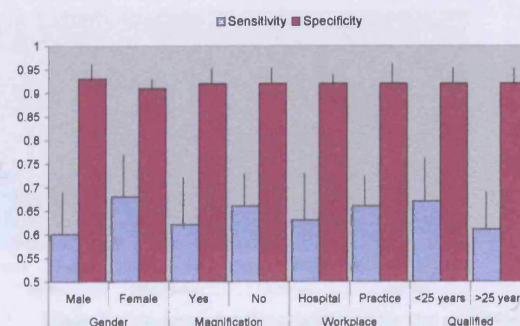


Figure 4.2b

Mean sensitivity and specificity values for the examiners during the pre-training simulated clinical phase

Table 4.3.3 Pre-training simulated clinical phase results evaluated by demographic parameters and expressed as mean, standard deviation and significance (Mann Whitney U test)

		Positive predictive value				Negative predictive value			Kappa agreement (self)			Kappa agreement to the gold standard		
Demographic	parameters	N	Mean	SD	p	Mean	SD	p	Mean	SD	p	Mean	SD	p
Gender	Male	8	0.81	0.1	0.24	0.88	0.04	0.06	0.68	0.09	0.92	0.61	0.07	0.25
	Female	8	0.75	0.13		0.81	0.07		0.69	0.05		0.55	0.06	
Magnification	Yes	9	0.79	0.09	0.16	0.83	0.07	0.22	0.67	0.09	0.75	0.57	0.09	0.60
	No	7	0.77	0.11		0.86	0.06		0.7	0.08		0.59	0.03	
Workplace	Hospital	10	0.79	0.07	0.51	0.83	0.07	0.48	0.7	0.05	0.23	0.57	0.07	0.66
	Practice	6	0.77	0.12		0.86	0.06		0.65	0.13		0.59	0.06	
Years qualified	<25 years	8	0.79	0.09	0.83	0.86	0.06	0.28	0.69	0.09	0.96	0.61	0.07	0.13
	>25 years	8	0.78	0.11		0.82	0.07		0.67	0.08		0.54	0.05	

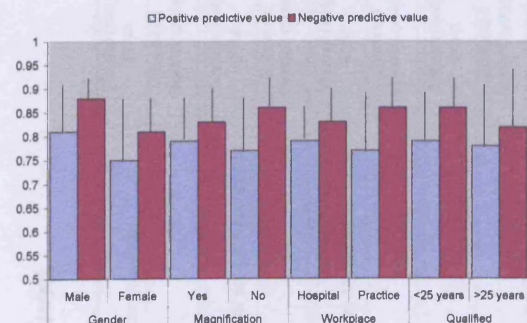


Figure 4.3 a Mean positive and negative predictive values illustrating the agreement between examiners clinical phase

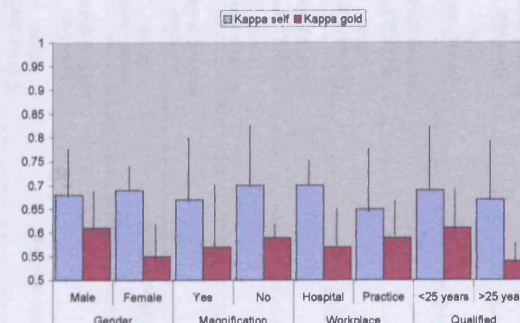


Figure 4.3b Mean kappa agreement calculations for the examiners during the pre-training simulated clinical phase during the pre-training simulated showing the agreement with themselves and the gold standard

Table 4.3.4 Differences between the trained (test) and untrained (control) group during the pre-training simulated clinical phase of the study

Variable examined		N	Mean	Std. deviation	Sig. (2 tailed)
Number of restorations scheduled for Replacement	Trained (test) group	8	34	7.93	0.46
	Untrained (control) group	8	36.3	7.78	
Dices Coincidence index for sound restorations	Trained (test) group	8	0.88	0.029	0.87
	Untrained (control) group	8	0.88	0.038	
Dices Coincidence index for replacing Restorations	Trained (test) group	8	0.69	0.052	0.39
	Untrained (control) group	8	0.71	0.048	
Sensitivity	Trained (test) group	8	0.64	0.088	0.67
	Untrained (control) group	8	0.64	0.089	
Specificity	Trained (test) group	8	0.92	0.031	0.53
	Untrained (control) group	8	0.93	0.027	
Positive Predictive Value	Trained (test) group	8	0.76	0.10	0.52
	Untrained (control) group	8	0.80	0.83	
Negative Predictive Value	Trained (test) group	8	0.85	0.067	0.56
	Untrained (control) group	8	0.83	0.72	
Kappa statistic – Agreement with Self	Trained (test) group	8	0.67	0.074	0.53
	Untrained (control) group	8	0.69	0.99	
Kappa Statistic - Agreement with Gold	Trained (test) group	8	0.57	0.067	0.53
	Untrained (control) group	8	0.585	0.72	
Washout period to exam 2	Trained (test) group	8	56.9	47.29	0.49
	Untrained (control) group	8	51.9	47.22	
Time taken for exam 1	Trained (test) group	8	39.1	8.54	0.75
	Untrained (control) group	8	38.9	4.16	
Time taken for exam 2	Trained (test) group	8	10.6	1.99	0.83
	Untrained (control) group	8	11.00	2.51	

Table 4.4.1 Raw data for the assessors showing the differences for the test (trained) group before and after training plus the control (untrained) group during the simulated clinical phases of the study

Test	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained	Untrained	Trained
γ^*	96	20	34	63	10	16	28	37	0.93	0.89	0.80	0.76	0.79	0.62	0.93	0.95	0.79	0.89	0.92	0.81	0.73	0.84	0.71	0.71
δ^*	94	19	45	56	9	17	41	21	0.85	0.96	0.66	0.81	0.56	0.91	0.93	0.91	0.82	0.73	0.78	0.97	0.65	0.94	0.52	0.82
ε^*	21	24	25	61	12	17	39	23	0.86	0.93	0.69	0.78	0.59	0.78	0.93	0.89	0.82	0.69	0.81	0.93	0.67	0.89	0.57	0.75
ζ^*	15	12	42	56	11	14	38	23	0.86	0.91	0.67	0.73	0.58	0.96	0.92	0.90	0.79	0.69	0.81	0.71	0.69	0.95	0.53	0.75
η^*	138	15	40	59	9	12	25	21	0.92	0.92	0.73	0.72	0.76	0.81	0.89	0.89	0.68	0.65	0.93	0.95	0.76	0.90	0.63	0.75
θ^*	14	20	51	65	14	17	46	44	0.86	0.81	0.71	0.64	0.57	0.92	0.97	0.93	0.93	0.85	0.76	0.71	0.65	0.90	0.57	0.58
μ^*	21	14	45	50	12	17	29	35	0.88	0.89	0.67	0.81	0.66	0.69	0.89	0.97	0.68	0.92	0.88	0.85	0.69	1.00	0.55	0.77
ν^*	56	21	31	64	8	12	26	36	0.88	0.87	0.63	0.71	0.65	0.61	0.87	0.94	0.61	0.85	0.89	0.81	0.51	0.67	0.51	0.69
Control																								
ξ	7		4		11		30		0.90		0.72		0.70		0.91		0.75		0.89		0.62		0.63	
$\omicron$	7		32		10		47		0.86		0.72		0.57		0.98		0.96		0.76		0.62		0.59	
π	18		40		15		37		0.88		0.71		0.62		0.93		0.82		0.83		0.85		0.59	
α	52		45		12		26		0.93		0.76		0.73		0.89		0.68		0.92		0.65		0.61	
β	2		43		9		44		0.84		0.67		0.55		0.94		0.86		0.76		0.77		0.52	
ι	14		35		8		29		0.92		0.77		0.76		0.93		0.79		0.92		0.52		0.69	
κ	117		38		9		43		0.82		0.62		0.51		0.91		0.79		0.75		0.72		0.45	
λ	108		38		14		34		0.89		0.71		0.65		0.92		0.79		0.86		0.75		0.60	

Table 4.4.2 Comparisons of the washout periods between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study.

Variable examined	Mean (days)	Std. deviation
First washout period (untrained control group)	51.88	47.22
First washout period (untrained test group)	56.88	47.29
Second washout period (trained test group)	18.13	4.05

	Test of significance
First washout period(untrained control group) v First washout period (untrained test group) ^M	0.49
First washout period (untrained control group) v Second washout period (trained test group) ^M	0.64
First washout period (untrained test group) v Second washout period (trained test group) ^W	0.058
	^M Mann-Whitney U test ^W Wilcoxon test

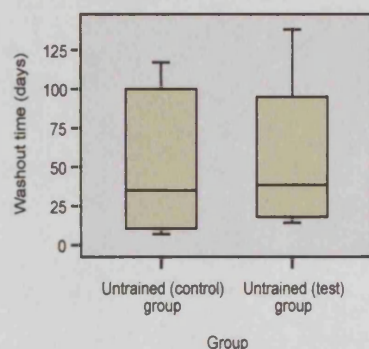


Figure 4.4.2a Box plots showing the differences in washout period between the untrained (control) group and the untrained (test) group during the pre-training simulated clinical phase

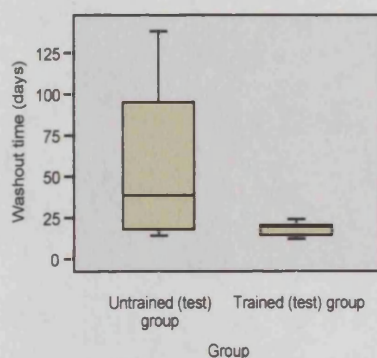


Figure 4.4.2b Box plots showing the differences in washout period between the untrained (test) group and the trained during the simulated clinical phases

Table 4.4.3 Comparisons of the full examination times taken between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study.

Variable examined	Mean	Std. deviation
Time taken for the full examination of the restorations (untrained control group)	38.88	4.16
Time taken for the full examination of the restorations (untrained test group)	39.13	8.54
Time taken for the full examination of the restorations (trained test group)	59.25	5.06

	Test of significance
Time taken for the full examination of the restorations (untrained control group) v Time taken for the full examination of the restorations (untrained test group) ^M	0.48
Time taken for the full examination of the restorations (untrained control group) v Time taken for the full examination of the restorations (trained test group) ^M	0.001*
Time taken for the full examination of the restorations (untrained test group) v Time taken for the full examination of the restorations (trained test group) ^W	0.012*

^M Mann-Whitney U test

^W Wilcoxon test

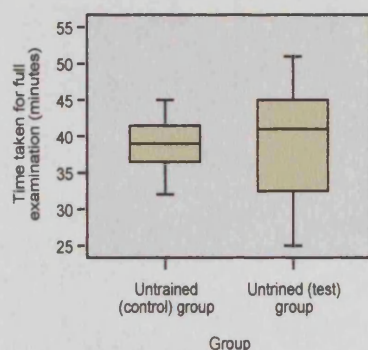


Figure 4.4.3a Box plots showing the differences in full examination times between the untrained (control) group and the untrained (test) group for the simulated clinical phases

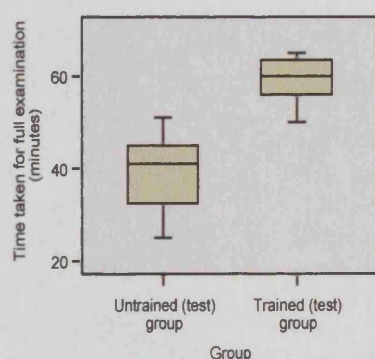


Figure 4.4.3b Box plots showing the differences in full examination times between the untrained (test) group and the trained (test) group for during the simulated clinical phases

Table 4.4.4 Comparisons of the time taken for the intra-examiner kappa statistic examinations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Variable examined	Mean	Std. deviation
Time taken to examine the restorations for the intra-examiner kappa statistic (untrained control group)	11.00	2.51
Time taken to examine the restorations for the intra-examiner kappa statistic (untrained test group)	10.63	1.99
Time taken to examine the restorations for the intra-examiner kappa statistic (trained test group)	15.25	2.25

	Test of significance
Time taken to examine the restorations for the intra-examiner kappa statistic (untrained control group) v Time taken to examine the restorations for the intra-examiner kappa statistic (untrained test group) ^M	0.42
Time taken to examine the restorations for the intra-examiner kappa statistic (untrained control group) v Time taken to examine the restorations for the intra-examiner kappa statistic (trained test group) ^M	0.008*
Time taken to examine the restorations for the intra-examiner kappa statistic (untrained test group) v Time taken to examine the restorations for the intra-examiner kappa statistic (trained test group) ^W	0.012*

^M Mann-Whitney U test ^W Wilcoxon test

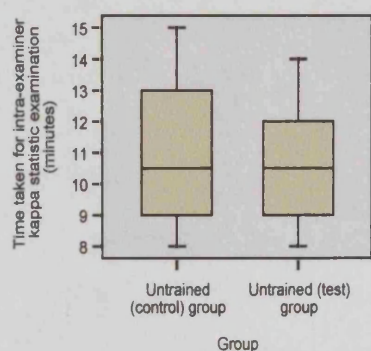


Figure 4.4a Box plots showing the differences in time taken for the intra-examiner kappa statistic examinations between the untrained (control) group and the untrained (test) group during the simulated clinical phases

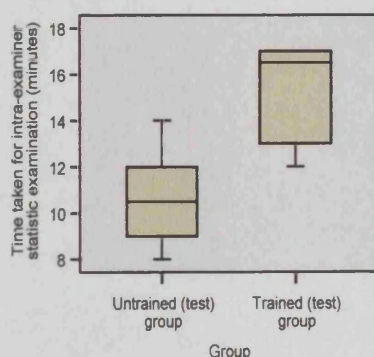


Figure 4.4b Box plots showing the differences in time taken for the intra-examiner kappa statistic examinations between the untrained (test) group and the trained (test) group during the simulated clinical phases

Table 4.4.5 Comparisons of the number of restorations scheduled for replacement between the untrained (control) group, untrained (test) group and the trained (test) group during the laboratory phases of the study

Variable examined	Mean	Std. deviation
Numbers of restorations scheduled for replacement (untrained control group)	36.25	7.78
Numbers of restorations scheduled for replacement (untrained test group)	34.00	7.93
Numbers of restorations scheduled for replacement (trained test group)	30.00	8.99

		Test of significance
Numbers of restorations scheduled for replacement (untrained control group)	v	0.64
Numbers of restorations scheduled for replacement (untrained test group) ^M		
Numbers of restorations scheduled for replacement (untrained control group)	v	0.12
Numbers of restorations scheduled for replacement (trained test group) ^M		
Numbers of restorations scheduled for replacement (untrained test group)	v	0.40
Numbers of restorations scheduled for replacement (trained test group) ^M		

^M Mann-Whitney U test ^W Wilcoxon test

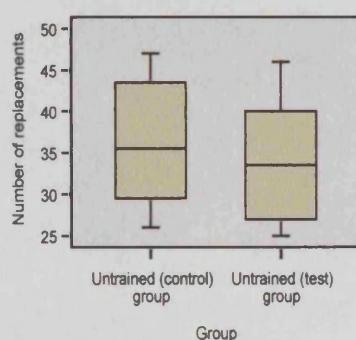


Figure 4.4.5a Box plots showing the differences in number of restorations scheduled for replacement between the untrained (control) group and the untrained (test) group during the simulated clinical phases of the study

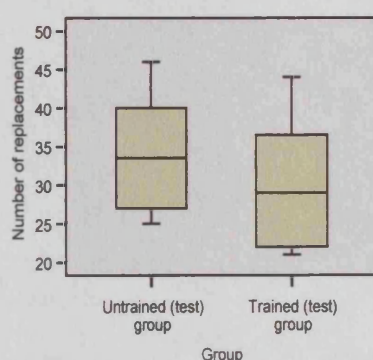


Figure 4.4.5b Box plots showing the differences in number of restorations scheduled for replacement between the untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Table 4.4.6 Comparison of the Dice's Coincidence Index for agreements that a restoration was sound between the untrained (control) group, untrained (test) group and the trained (test) group during the laboratory phases of the study

Variable examined	Mean	Std. deviation
Dice's Coincidence Index for agreements that restorations were sound (untrained control group)	0.88	0.038
Dice's Coincidence Index for agreements that restorations were sound (untrained test group)	0.88	0.029
Dice's Coincidence Index for agreements that restorations were sound (trained test group)	0.89	0.045

Test of significance	
Dice's Coincidence Index for agreements that restorations were sound (untrained control group) v Dice's Coincidence Index for agreements that restorations were sound (untrained test group) ^M	0.56
Dice's Coincidence Index for agreements that restorations were sound (untrained control group) v Dice's Coincidence Index for agreements that restorations were sound (trained test group) ^M	0.38
Dice's Coincidence Index for agreements that restorations were sound (untrained test group) v Dice's Coincidence Index for agreements that restorations were sound (trained test group) ^W	0.39

^M Mann-Whitney U test ^W Wilcoxon test

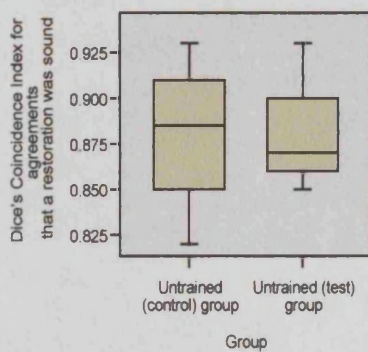


Figure 4.4.6a Box plots showing the differences in full examination times between the untrained (control) group and the untrained (test) group for the simulated clinical phases

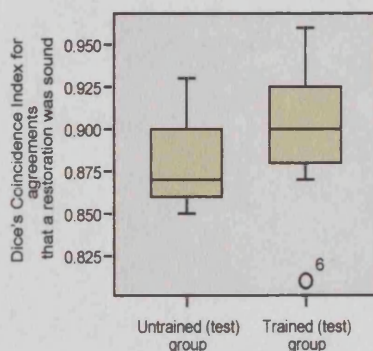


Figure 4.4.6b Box plots showing the differences in full examination times between the untrained (test) group and the trained (test) group for the initial and simulated clinical phases

Table 4.4.7 Comparison of the Dice's Coincidence Index for agreements that a restoration required replacement between the untrained (control) group, untrained (test) group and the trained (test) group during the laboratory phases of the study

Variable examined	Mean	Std. deviation
Dice's Coincidence Index for agreements that restorations needed replacing (untrained control group)	0.71	0.048
Dice's Coincidence Index for agreements that restorations needed replacing (untrained test group) ^M	0.69	0.052
Dice's Coincidence Index for agreements that restorations needed replacing (trained test group)	0.75	0.057

	Test of significance
Dice's Coincidence Index for agreements that restorations needed replacing (untrained control group) v Dice's Coincidence Index for agreements that restorations needed replacing (untrained test group) ^M	0.162
Dice's Coincidence Index for agreements that restorations needed replacing (untrained control group) v Dice's Coincidence Index for agreements that restorations needed replacing (untrained test group) ^M	0.19
Dice's Coincidence Index for agreements that restorations needed replacing (untrained test group) v Dice's Coincidence Index for agreements that restorations needed replacing (trained test group) ^W	0.12

^M Mann-Whitney U test ^W Wilcoxon test

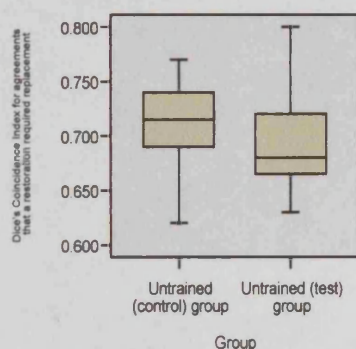


Figure 4.4.7a Box plots showing the differences in Dice's Coincidence Index for agreements that a restoration required replacement between the untrained (control) group and the untrained (test) group during the simulated clinical phases

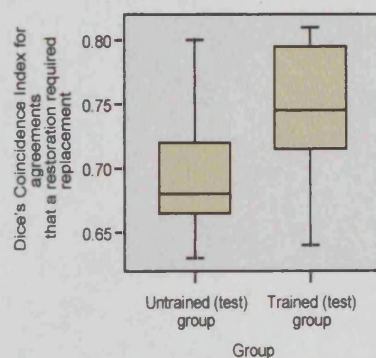


Figure 4.4.7b Box plots showing the differences in Dices's Coincidence Index for agreements that a restoration required replacement between the untrained (test) group and the trained (test) during the simulated clinical phases

Table 4.4.8 Comparisons of the sensitivity calculations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Variable examined	Mean	Std. deviation
Sensitivity (untrained control group)	0.64	0.089
Sensitivity (untrained test group)	0.64	0.088
Sensitivity (trained test group)	0.78	0.14

	Test of significance
Sensitivity (untrained control group) v Sensitivity (untrained test group) ^M	1.00
Sensitivity (untrained control group) v Sensitivity (trained test group) ^M	0.021*
Sensitivity (untrained test group) v Sensitivity (trained test group) ^{+W}	0.093

^M Mann-Whitney U test

^W Wilcoxon test

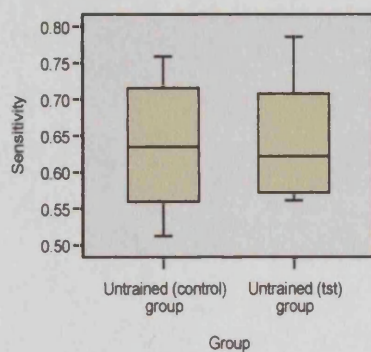


Figure 4.4.8a Box plots showing the differences in Sensitivity between the untrained (control) group and the untrained (test) group during the simulated clinical phases

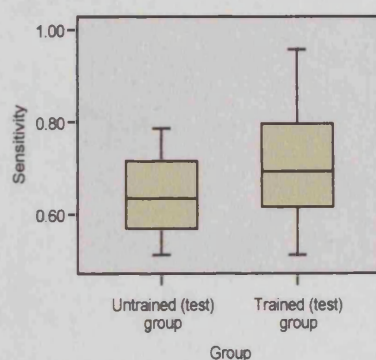


Figure 4.4.8b Box plots showing the differences in Sensitivity between the untrained (test) group and the trained (test) group during the simulated clinical phases

Table 4.4.9 Comparison of the specificity calculations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Variable examined	Mean	Std. deviation
Specificity (untrained control group)	0.93	0.027
Specificity (untrained test group)	0.92	0.031
Specificity (trained test group)	0.92	0.029

Test of significance	
Specificity (untrained control group) v Specificity (untrained test group) ^M	0.42
Specificity (untrained control group) v Specificity (trained test group) ^M	0.49
Specificity (untrained test group) v Specificity (trained test group) ^w	0.89

^M Mann-Whitney U test ^w Wilcoxon test

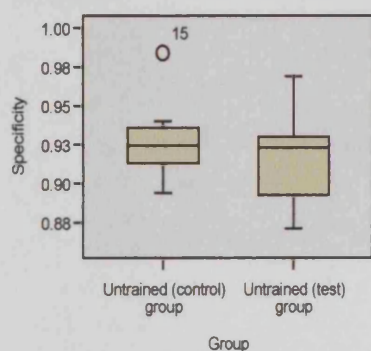


Figure 4.4.9a Box plots showing the differences in Specificity between the untrained (control) group and the untrained (test) group during the simulated clinical phases of the study

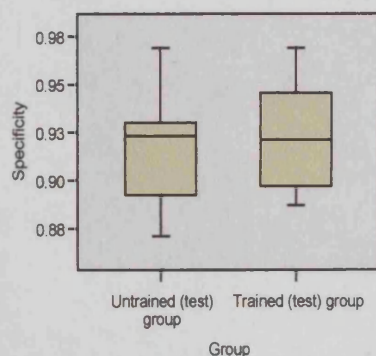


Figure 4.4.9b Box plots showing the differences in Specificity between the untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Table 4.4.10 Comparison of the positive predictive value calculations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Variable examined	Mean	Std. deviation
Positive Predictive Value (untrained control group)	0.80	0.083
Positive Predictive Value (untrained test group)	0.76	0.10
Positive Predictive Value (trained test group)	0.78	0.10

Test of significance	
Positive Predictive Value (untrained control group) v Positive Predictive Value (untrained test group)	0.52
Positive Predictive Value (untrained control group) v Positive Predictive Value (trained test group)	0.49
Positive Predictive Value (untrained test group) v Positive Predictive Value (trained test group)	0.78

^M Mann-Whitney U test ^W Wilcoxon test

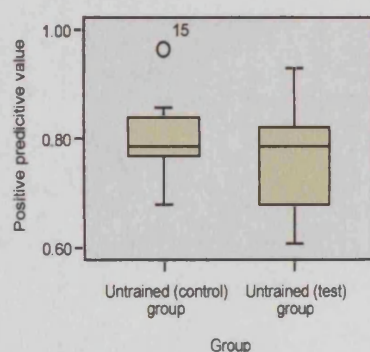


Figure 4.4.10a Box plots showing the differences in Positive predictive values between the untrained (control) group and the untrained (test) group during the simulated clinical phases of the study

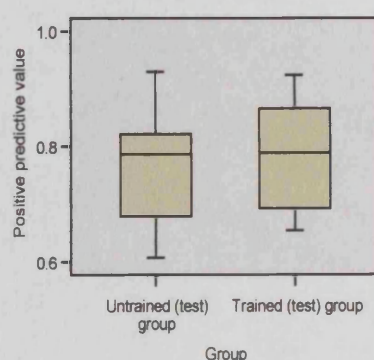


Figure 4.4.10b Box plots showing the differences in Positive predictive values between the untrained (control) group and the trained (test) group during the simulated clinical phases of the study

Table 4.4.11 Comparison of the negative predictive value calculations between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Variable examined	Mean	Std. deviation
Negative Predictive Value (untrained control group)	0.83	0.072
Negative Predictive Value (untrained test group)	0.85	0.067
Negative Predictive Value (trained test group)	0.84	0.10

Test of significance	
Negative Predictive Value (untrained control group) v Negative Predictive Value (untrained test group) ^M	0.86
Negative Predictive Value (untrained control group) v Negative Predictive Value (trained test group)	0.73
Negative Predictive Value (untrained test group) v Negative Predictive Value (trained test group) ^W	0.78
^M Mann-Whitney U test ^W Wilcoxon test	

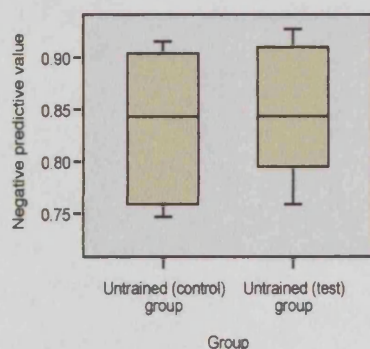


Figure 4.4.11a Box plots showing the differences in Negative predictive values between the untrained (control) group and the untrained (test) group during the simulated clinical phases of the study

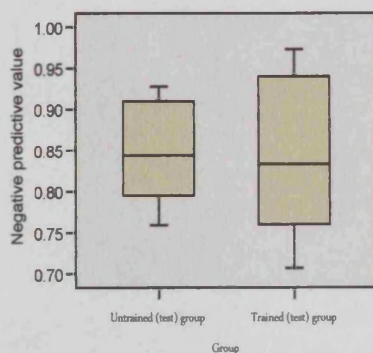


Figure 4.4.11b Box plots showing the differences in Negative predictive values between the untrained (control) group and the trained (test) group during the simulated clinical phases of the study

Table 4.4.12 Comparison of the intra-examiner kappa statistic for variability between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Variable examined	Mean	Std. deviation
Kappa statistic for intra-examiner variability (untrained control group)	0.69	0.099
Kappa statistic for intra-examiner variability (untrained test group)	0.67	0.074
Kappa statistic for intra-examiner variability (trained test group)	0.89	0.10

	Test of significance
Kappa statistic for intra-examiner variability (untrained control group) v Kappa statistic for intra-examiner variability(untrained test group) ^M	0.42
Kappa statistic for intra-examiner variability (untrained control group) v Kappa statistic for intra-examiner variability(trained test group) ^M	0.008*
Kappa statistic for intra-examiner variability (un trained test group) v Kappa statistic for intra-examiner variability (trained test group) ^{*w}	0.012*

^M Mann-Whitney U test ^w Wilcoxon test

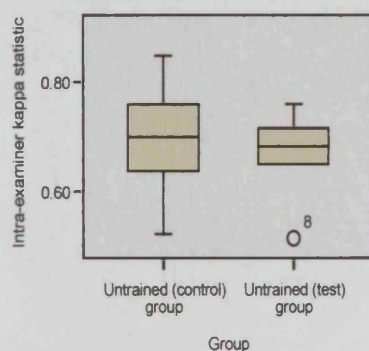


Figure 4.4.12a Box plots showing the differences in intra-examiner kappa statistic between the untrained (control) group and the untrained (test) group during the simulated clinical phases of the study

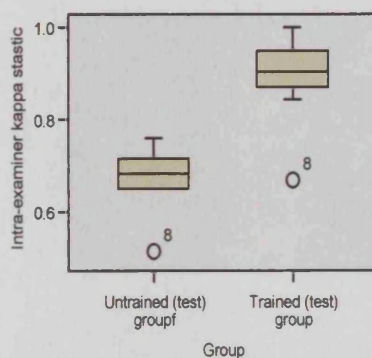


Figure 4.4.12b Box plots showing the differences in intra-examiner kappa statistic between the untrained (control) group and the trained (test) group during the simulated clinical phases of the study

Table 4.4.13 Comparison of the inter-examiner kappa statistic for variability between the untrained (control) group, untrained (test) group and the trained (test) group during the simulated clinical phases of the study

Variable examined	Mean	Std. deviation
Kappa statistic for inter-examiner agreement (untrained control group)	0.59	0.072
Kappa statistic for inter-examiner agreement (untrained test group)	0.57	0.062
Kappa statistic for inter-examiner agreement (trained test group)	0.73	0.073

Test of significance	
Kappa statistic for inter-examiner agreement (untrained control group) v Kappa statistic for inter-examiner agreement (untrained test group) ^M	0.25
Kappa statistic for inter-examiner agreement (untrained control group) v Kappa statistic for inter-examiner agreement (trained test group) ^M	0.0015*
Kappa statistic for inter-examiner agreement untrained test group v Kappa statistic for inter-examiner agreement trained test group) ^W	0.017*

^M Mann-Whitney U test

^W Wilcoxon test

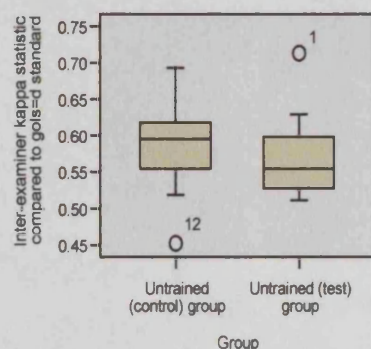


Figure 4.4.13a Box plots showing the differences in inter-examiner kappa statistic between the untrained (control) group and the untrained (test) group during the simulated clinical phases of the study

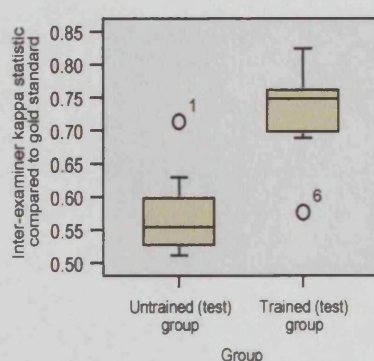


Figure 4.4.13b Box plots showing the differences in inter-examiner kappa statistic between the untrained (control) group and the trained (test) group during the simulated clinical phases of the study

Table 4.5.1 Summary data for patients taking part in the clinical phase

No. of patients	Gender	Age range (years)	Mean age (years)	Std. Dev
9	Female	19 - 54	30.6	$\pm$ 12.00

Total number of restorations	64	
Total number of amalgam restorations	34	(1 subsequently lost due to patient receiving dental treatment)
Total number of resin restorations	30	
Total number of restorations in need of re-restoring	5	(3 amalgam, 2 resin)

Table 4.5.2 Raw data, evaluations and subsequent statistical calculations for the clinical phase

Assessor	Washout (days)	Time taken to carry out the examination of the full cohort of patients (mins)	Time taken to carry out the examination of the patients during the kappa statistic determination (mins)	Number of restorations scheduled for replacement	Dices' Coincidence Statistic for agreement that a restoration was sound	Dices' Coincidence Statistic for agreement that a restoration required replacement	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Intra-examiner kappa value	Inter-examiner kappa value when compared to the gold standard
α	7	39	10	12	0.89	0.38	0.25	0.98	0.75	0.82	0.79	0.77
β	15	35	11	11	0.87	0.38	0.25	0.97	0.75	0.79	0.85	0.78
γ^*	7	33	11	10	0.88	0.46	0.30	0.97	0.75	0.83	0.87	0.86
δ^*	18	26	nr	1	0.95	0.00	0.00	0.93	0.0	0.98	1.00	0.87
ε^*	10	31	8	4	0.95	0.25	0.25	0.94	0.25	0.94	0.90	0.83
ζ^*	14	24	8	7	0.91	0.36	0.29	0.95	0.5	0.88	0.92	0.84
η^*	10	24	8	5	0.95	0.44	0.40	0.96	0.40	0.94	0.92	0.87
κ	21	43	17	10	0.93	0.57	0.40	1.00	1.00	0.88	0.86	0.73
λ	24	38	14	12	0.87	0.25	0.17	0.65	0.5	0.79	0.80	0.73
μ^*	8	29	9	6	0.96	0.60	0.50	0.98	0.75	0.94	0.95	0.90
ν^*	27	28	10	9	0.92	0.50	0.33	0.97	0.75	0.84	0.81	0.84
ξ	21	34	12	3	0.95	0.29	0.33	0.94	0.25	0.96	0.93	0.76
$\omicron$	10	32	10	11	0.83	0.13	0.091	0.93	0.25	0.77	0.77	0.71
π	10	36	17	9	0.93	0.46	0.33	0.98	0.75	0.89	0.95	0.81

* represents trained assessors

nr this assessor was not asked to carry out a measurement for intra-examiner variability as they only identified one restoration for replacement and was able to recall this clearly.

Table 4.5.3 Means, standard deviations and results of the statistical comparisons of the means (Mann-Whitney U test) for the untrained (control) group and the trained (test) group for the clinical phase of the project

Variable examined		Mean	Std. deviation	Sig. (2 tailed)
Number of restorations scheduled for replacement	Untrained (control) group	9.71	3.15	0.034*
	Trained (test) group	6.00	3.06	
Washout period between initial exam and exam for calculating the intra-examiner kappa statistic	Untrained (control) group	15.4	6.66	0.52
	Trained (test) group	13.4	7.07	
Time taken to examine the full cohort of patients and restorations	Untrained (control) group	36.7	3.64	0.003*
	Trained (test) group	27.9	3.44	
Time taken to complete the intra-examiner kappa statistic examination	Untrained (control) group	13.0	3.06	0.011*
	Trained (test) group	9.00	1.265	
Dice's Coincidence Index for sound restorations	Untrained (control) group	0.89	0.029	0.092
	Trained (test) group	0.93	0.042	
Dice's Coincidence Index for replace restorations	Untrained (control) group	0.35	0.14	0.65
	Trained (test) group	0.37	0.19	
Sensitivity	Untrained (control) group	0.56	0.22	0.44
	Trained (test) group	0.60	0.33	
Specificity	Untrained (control) group	0.91	0.054	0.37
	Trained (test) group	0.94	0.061	
Positive predictive value	Untrained (control) group	0.59	0.20	0.95
	Trained (test) group	0.59	0.37	
Negative predictive value	Untrained (control) group	0.76	0.34	0.13
	Trained (test) group	0.94	0.061	
Intra-examiner kappa statistic	Untrained (control) group	0.88	0.045	0.14
	Trained (test) group	0.92	0.067	
Inter-examiner kappa statistic Compared to the gold standard	Untrained (control) group	0.79	0.061	0.002*
	Trained (test) group	0.85	0.27	

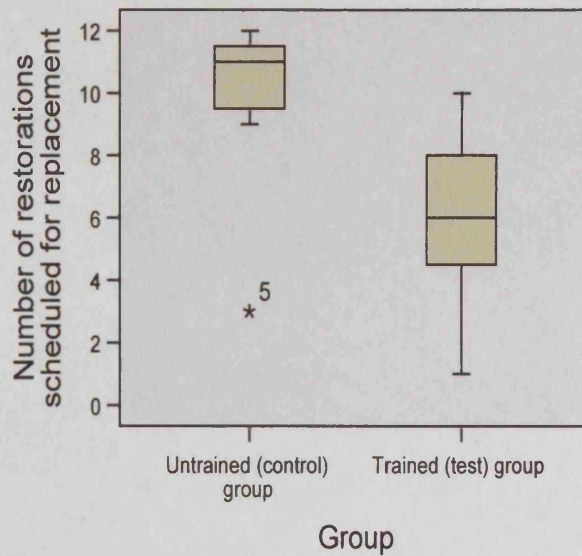


Figure 4.5.3a A box plot showing the difference between the untrained (control) group and the trained (test) group with respect to the number of restorations scheduled for replacement in the clinical phase of the project ($p = 0.034$)

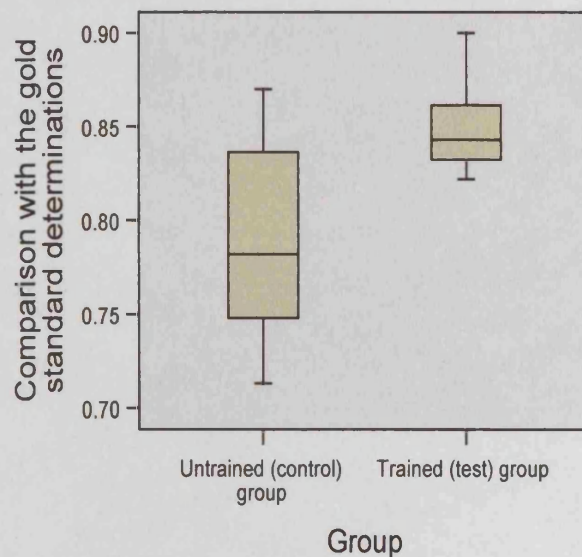


Figure 4.5.3b A box plot showing the difference between the untrained (control) group and the trained (test) group with respect to agreeing with the gold standard set for the clinical phase of the project ($p = 0.002$)

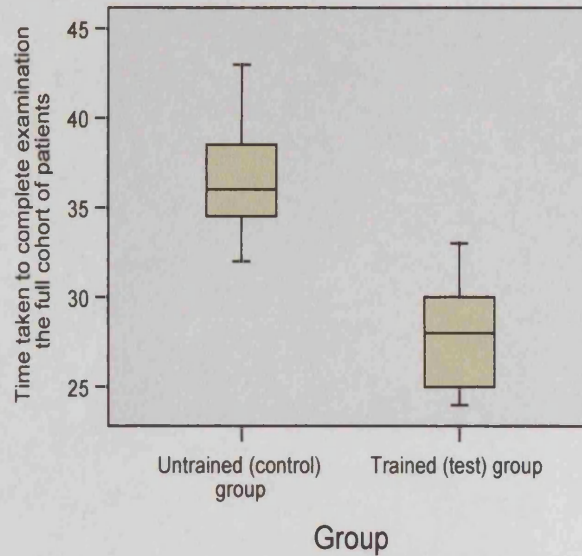


Figure 4.5.3c A box plot showing the difference between the untrained (control) group and the trained (test) group with respect to the time taken to examine the full cohort of patients during the clinical phase of the project ($p = 0.003$)

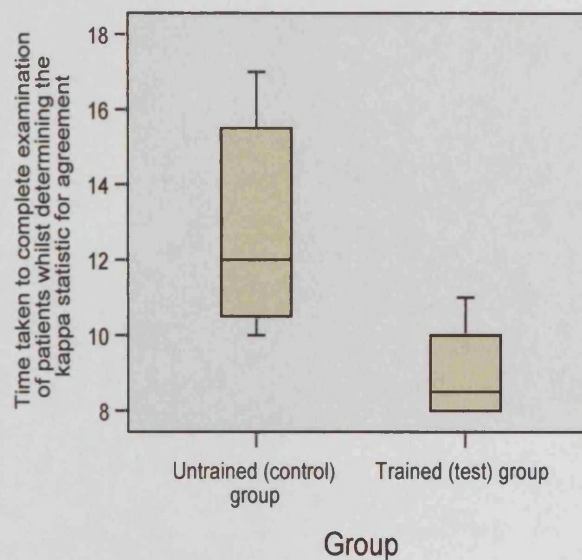


Figure 4.5.3d A box plot showing the difference between the untrained (control) group and the trained (test) group with respect to the time taken to examine patients during the intra-examiner kappa statistic examinations of patients in the clinical phase of the project ($p = 0.011$)

Table 4.6.1 Responses to the evaluation survey**Section A**

Question	Very easy	Easy	Difficult	Very difficult
Overall, how would you rate the applicability of the evaluation criteria to clinical practice?	2	4	2	0
How would you rate the applicability of the colour component of the evaluation criteria to clinical practice?	2	5	1	0
How would you rate the applicability of the anatomical form component of the evaluation criteria to clinical practice?	2	5	1	0
How would you rate the applicability of the marginal integrity component of the evaluation criteria to clinical practice?	3	4	1	0
How would you rate the applicability of the caries component of the evaluation criteria to clinical practice?	3	3	2	0

Section B

Question	Yes	No	No reply
Do you feel that the use of the evaluation criteria makes it easier for you to decide if a restoration needs replacing?	7	1	
Do you feel your reliability and consistency with respect to restoration replacement need is improved through the application of the evaluation criteria?	6	1	1
Do you think that these evaluation criteria have a role in clinical decision making?	8	0	
Will you continue to use the evaluation criteria in your everyday working practice?	7	1	
Has taking part in this project altered your clinical practice when it comes to decision making with respect to restoration replacement?	5	3	

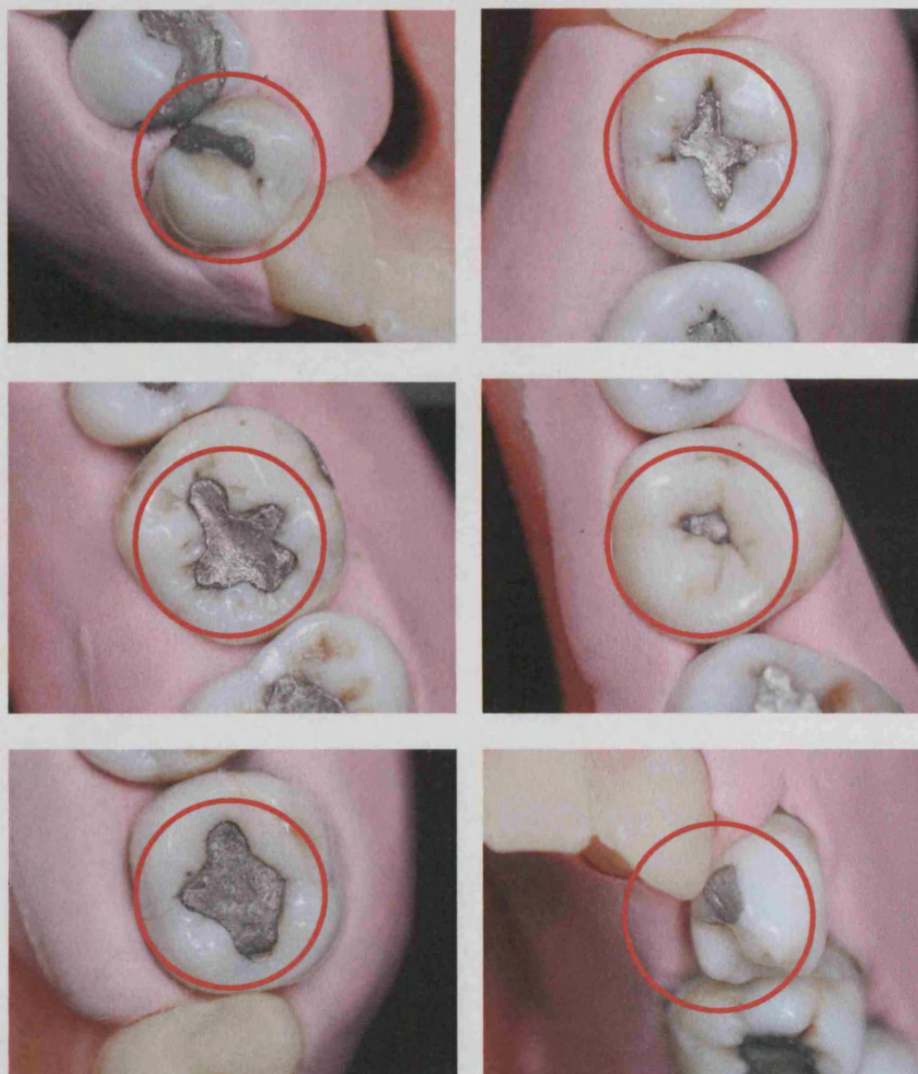
Table 4.6.2 Free comments generated by the evaluation questionnaire. The number of times a comment was made is indicated in brackets

-
- **I feel that this method of assessing restorations is extremely valuable. I now use it nearly every day in practice (2)**
-
- **I found the assessment method hard to use at first but things progressed as I became used to it. I think I am more reliable in my decision making now (2)**
 - **This type of assessment methodology would lend itself to an undergraduate training programme in order to help them with decision making (4)**
 - **I still find caries diagnosis at the edge of restorations hard (1)**
 - **When can I undertake training? I would like too! (3)**
 - **Is there any evidence to link the assessment criteria with restoration longevity? (2)**
-

FIGURES FOR CHAPTER 4

RESULTS

Figure 4.1 Examples of restorations where the gold standard operators disagreed



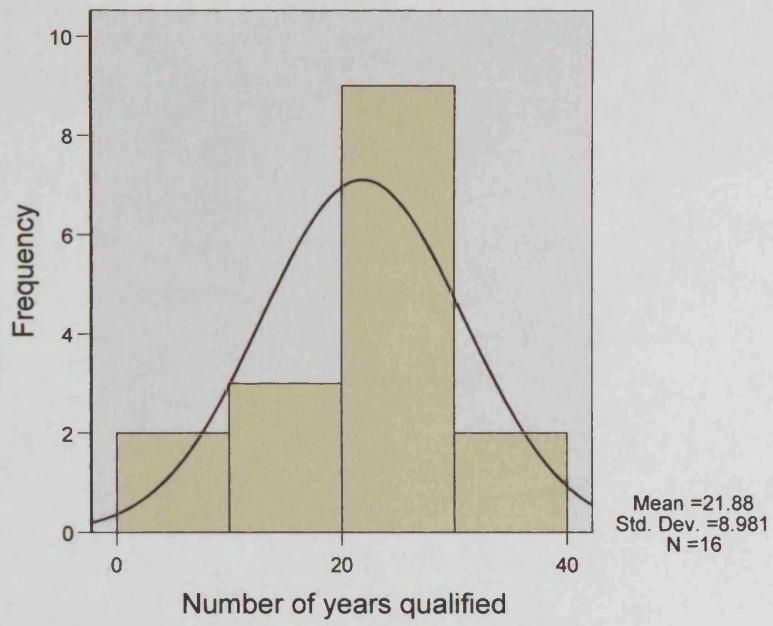


Figure 4.2 Graph showing the number of assessors by the number of years qualified

CHAPTER 5

DISCUSSION

5 Discussion

As outlined in the materials and methods (Section 3.1) the project had a number of clearly defined phases that depended on identifiable targets within them; these are discussed chronologically before considering some more general points highlighted by the research.

5.1 Pre-simulated clinical phase

5.1a Selection of teeth and construction of models

In dentistry, new techniques and methods are generally evaluated before being introduced to the clinical situation and these evaluations are commonly made on models using extracted teeth. Traditionally, evaluations have been made in an artificial way with teeth being embedded in dental stone silicone or plasticine (Maupomé, 1998; Rudolphy *et al*, 1995; Merrett and Elderton, 1984) but more recently the practice of setting natural teeth into artificial jaws and examining them in a phantom head under simulated clinical conditions has been employed (Erten *et al*, 2005; Forgie *et al*, 2002; Deery *et al*, 1995; McAndrew and Longbottom, 1993). The use of models in the simulated clinical phases kept the experimental set up simple, allowed flexibility and gave scope to investigate logistical problems in the research without inconveniencing and unnecessarily involving patients in the early stages of the project's development and although it produced initially an *in vitro* measure of diagnostic ability of the operators a number of steps were taken in order to replicate the clinical situation as closely as possible; models were mounted in phantom heads, pink silicone was used to simulate gingival tissue, the teeth were examined under clinical conditions using commonly available tools and real human teeth restored by a number of unknown dentists of differing clinical

ability were used which allowed the assessment of a range of restorations of varying quality – obviously good to obviously poor. It was felt desirable in this research to replicate as nearly as possible “real life” and this is why teeth were mounted and viewed as described; there is no research to show that diagnostic accuracy is increased by being able to freely pick-up, turn and manipulate an embedded tooth this is also unrealistic of what can be achieved clinically. It would indeed be interesting to determine if mounting teeth in a manner similar to that used in stand alone or clinical type models influences diagnostic ability amongst practitioners.

The use of natural human teeth for caries assessment and treatment decisions with respect to restoration replacement is not new (Maupomé 1998; Ketley and Holt, 1993; Lussi, 1993; Kay *et al*, 1988; Merrett and Elderton, 1984) and previous research using clinical models to investigate caries diagnosis have faced similar dilemmas to those experienced here when selecting extracted teeth to construct models (Forgie *et al*, 2002; Hintze *et al*, 1998; Ekstrand *et al*, 1997; Deery *et al*, 1995; Ricketts *et al*, 1995). With specific reference to the teeth and restorations evaluated in the study, it was felt necessary for them to be representative of that which would be normally encountered clinically. It is important to remember that in practice clinical decisions on restoration replacement are generally easier to make if they are associated with obvious defects but that restorations with more limited defects give rise to uncertainty and variability in treatment decisions (Kay and Nuttall, 1994; Bader and Shugars, 1993; Bader and Shugars, 1992; Anusavise, 1989; Anusavise, 1985; Elderton, 1983). It was, therefore, important to ensure that the restorations used presented a reasonable degree of diagnostic challenge. It has been highlighted that

restoration replacement rates vary considerably within populations (Table 2.2) and it is clear from clinical experience that the “global average” of around 50% is not always encountered. Deery *et al*’s (1995) work (carried out in the UK) suggests that a caries prevalence of 28% is not only achievable and workable in research protocols but also realistic of what could be encountered clinically; this translates to between 1 in 3 or 1 in 4 teeth being carious. In this study the clinical variation in restoration replacement rate was reflected in a variable restoration replacement rate in the models which ranged from 11 to 60% (median 22%, mean 28%). It was believed that the levels of diagnostic uncertainty in the samples would be sufficient to prevent restoration recognition and recall to be shown by the volunteer assessors taking part in the different phases of the study. Indeed, this was borne out throughout the research as only on one occasion did any of the assessors recollect clearly the decisions they had made at a previous assessment. This occurred during the clinical phase where one assessor nominated only one restoration for replacement; the operator being able to clearly recall this after the washout period. This finding highlights that if a sample contains clearly identifiable or small numbers of “failing” restorations i.e. a sample in which lacks diagnostic challenge then difficulties maybe encountered if an assessor prescribes very conservatively. The only realistic way of avoiding this would be to use a larger sample or increase the number of obviously deficient restorations. The fact that “total recall” only occurred on one occasion may give credence to the feeling that the degree of restoration inadequacy was indeed sufficient and, fortunately, avoided the need to establish a larger more complex study.

As imagined, a significant number of extracted teeth had to be examined to produce a suitable sample. A common pool of extracted teeth may not be the best

sample to use as the teeth have generally been removed for different reasons and this can create difficulties; heavily restored teeth extracted because of pulpal or periapical pathology are often inherently unsuitable for evaluation in research project such as this one, teeth extracted for periodontal, orthodontic or elective (impacted teeth) reasons were equally unsuitable if they had no restoration at all. Today, the pooling of restorations and teeth themselves for research purposes has to follow the guidelines and statutes laid down (Human Tissue Act 2004, www.opsi.gov.uk/acts/acts2004/20040030.htm). In this study the teeth had been collected before the statute came into force.

The collected teeth were mainly premolars and molars with the most common restorative material being amalgam. This being representative of that placed clinically over time. There has however been a noticeable trend towards the increased use of composite resin in posterior teeth; it being estimated at around 30% in load bearing areas from research carried out in Ireland, the USA and the UK (Lynch *et al*, 2007).

Within the sample of teeth evaluated very few filled anterior teeth or resin based restorations were collected and a decision had to be made to limit the study to premolar and molar teeth in order that the research could continue; this was not unexpected as traditionally such extracted teeth have been used for endodontic training purposes in Cardiff and were unavailable. It was also known that anterior teeth generally receive fewer restorations (Tyas, 2005). The limited supply of anterior filled teeth may suggest that a protocol similar to the simulated clinical one used here may not be practical and that a more clinically orientated protocol along the lines of that carried out may be necessary.

The effects that the Human Tissue Act and the changes in restorative material use on obtaining a sample of teeth like that used in this study is as yet undetermined. Ultimately as the assessment criteria are applicable to all teeth with restorations the pragmatic decision to limit it to the tooth types available should be minimal, although this may merit further investigation.

The use of mainly amalgam restorations (107 of the 111) reduced the opportunity for the volunteer assessors to apply the colour component of the USPHS criteria. This was not felt to be a major problem in the research as a number of tooth coloured restorations had been included. In this study the gold standard defined replacement rates for the tooth coloured restorations was 33% (1 of 4) and 29% (31 of 107) for the amalgam restorations. Ideally it would have been desirable for more resin based restorations to be included in the simulated clinical phase; it was unfortunate but a harsh reality that the sample collected failed to provide the researchers with such an opportunity. Future studies will need to look at the effect of greater numbers of tooth coloured materials in the processes evaluated.

Single surface restorations were used, preferentially, in order to reduce examination difficulty when applying the assessment criteria to the less accessible interproximal areas. However, 50 of the 111 restorations involved more than one surface of the tooth. The impact that extensive restorations have on the applicability of using USPHS is something which would merit further investigation as does the inclusion of appropriate radiographs which can be used to supplement interproximal examinations.

The “pre-use” storage conditions were an initial concern as the ages of the teeth were indeterminable and their robustness during the pre-clinical phase of the

study potentially unpredictable. To reduce tooth fracture or restoration loss and to stay within the boundaries of the study with respect to diagnostic accuracy no teeth were accepted into the study that were severely decayed (cavitations that entered the pulp space), heavily restored (more than two surfaces) or grossly discoloured (where it was difficult to distinguish between tooth and restoration) by means of intrinsic or extrinsic stain; discolouration at the margin of restorations has been suggested as a complicating factor when making a diagnosis of recurrent caries (Ettinger, 1999). The teeth and restorations used stood up well to the multiple examinations throughout the study and this is testimony to the storage conditions used between examinations; unfortunately the storage conditions are not always fully described in other research projects and this limits comparisons with this research, although Deery *et al*'s work (1995) and the previous experiences of the principal researcher (McAndrew and Longbottom, 1994, 1993) realise the importance of keeping models fully hydrated in order to prevent desiccation. In the originally collected sample, one restoration was dislodged on examination and one tooth fractured. Both these failures took place within the first few examinations of the pre-training simulated clinical phase and as they had only been evaluated by three of the untrained volunteers their results were discarded and not included in the analyses. The effect of including the excluded groups of teeth into a similar study on the results and the performance of models could be determined in a later project. Unfortunately research projects similar to this one where teeth were anatomically set-up and mounted fail to fully indicate their inclusion and exclusion criteria or indeed failure rates within the models which again makes comparisons between projects impossible (Erten *et al*, 1995; Ermis and Aydin, 2004; Forgie *et al*, 2002).

5.1b Determining the gold standard for restoration replacement

Fyfe *et al* (2000) suggest that results or standard setting that are achieved through lack of calibration prevents the meaningful comparison of datasets. This suggests that examiner agreement is a pre-requisite in research where a gold standard can be set. The gold standard with respect to restoration replacement in this research was from the independent and then collaborative evaluations of two experienced researchers and this an accepted practice with respect to getting agreement between examiners in restoration replacement (Elderton 1977a; Ryge *et al*, 1974; Eames *et al*, 1974; Osborne and Gale, 1974a; Osborne *et al*, 1973; Phillips *et al*, 1973; Hedegård, 1955). It is however a clinical approach that deviates from that used in laboratory studies to determine presence or absence of parameters such as caries in extracted teeth; normally obtained through the examination of sectioned teeth (Erten *et al*, 2005; Forgie *et al*, 2002; Merrett and Elderton, 1984). As it was clearly not possible to section the teeth used in the clinical phase - a practitioner cannot sacrifice a tooth just to confirm a diagnosis! - it was deemed to be more appropriate to apply one consistent approach for setting the gold standard throughout all phases of the study, namely a clinically appropriate examination. Two examiners were chosen because agreement between examiners becomes more difficult as the number of examiners increases (Poulsen *et al*, 1980; Gröndahl, 1979; Davies and Caldell, 1963; Slack *et al*, 1958). Equally, such an approach provided the optimal conditions for achieving agreement and consistency during clinical examinations employing an assessment methodology that was well understood, clearly defined and practiced after appropriate training (Garbuz, 2002; Elderton 1977a).

It has to be noted that neither of the gold standard assessors used in this study had completed a formal or recognised training programme in the use of the USPHS criteria. However, their familiarity with the system was achieved from a working knowledge and its practical use in a field trial (BC). In this research the principal investigator (RM) was “trained and mentored ” by the other gold standard operator through the delivery of written material, one-to-one explanation of the characteristics of USPHS and the verification of understanding through validated assessments of the restorations included for analysis in the simulated and clinical phases of the study. One could argue that two different gold standard assessors may have identified different restorations for replacement. However, this is conjecture as it is equally possible that two other gold standard assessors with equal insight into the rudiments of USPHS would have identified the same restorations as needing replacement. A similar study with two different groups of assessors could confirm this.

In order to reduce examination variables a consistent approach to participant examination was taken at all stages of the study with standard instrumentation and identical clinical conditions being provided. Additionally, assessors were asked to evaluate the restorations on an individual basis and to assume the patient’s dentition was problem free. A similar approach has been taken by other researchers (Mills and Hollis, 1997) in an attempt to reduce examination confounders.

Although both examiners identified similar numbers of restorations as requiring replacement (Table 4.1.2) on the first examination, total agreement was only achieved for 15 of these (17 of the 32 had some area of disagreement which varied from 1 to 4 fields) and this is illustrated in Table 4.1.1e and Figure 5.1.

Table 4.1.1c also shows that overall differences in the assessment criteria were noted for 84 of the 111 restorations; disagreement in 1 field being noted for 28, 2 disagreements in 2 fields for 20, disagreements in 3 fields for 27 and disagreements in 4 fields for 9. Despite these findings it was noted that the gold standard assessors agreed that the majority of restorations were functional but only differed on how far along the scale of failure they sat. Such variation in treatment decisions is not new (Boyd, 1989; Kennon, 1989; Hocott, 1984; Elderton and Nuttall, 1983; Milgrom *et al*, 1981; Natkin and Guild, 1967; Abramowitz, 1966; Gruebbel, 1950). (For ease of reference a graphical display of these findings is given in Figure 5.1). The assessment criteria that produced the greatest differences in evaluation were noted in to be in marginal discolouration (64 of 111) and marginal adaptation fields (52 of 111) with anatomical form field (38 of 111) and caries assessment fields following (20 of 111). However, it was noted that on no occasion did the gold standard assessor's opinion differ by more than one scoring criterion between each other. The percentage for unequivocal agreement between the gold standard assessors was 65 percent based on the first examination ($197/555 \times 100$) and just over 66% ($187/555 \times 100$) on the second examination. This compares similarly with Robertello and Pinks study (1997) that examined the effect of a training programme (based on a rating scale published by Charbeneau in 1975) on the reliability of examiners evaluating amalgam restorations and reported a mean interexaminer agreement of 61%.

On the second examination, the number of restorations scheduled for replacement dropped - 28 of 111 (BC) and 27 of 111 (RM) (Table 4.1.2). However, of these restorations only one restoration was identified where disagreement was noted between the gold standard assessors (Restoration No.

20793L6O, Figure 5.2). It was noted that no restoration scheduled for replacement at the second examination was not included by at least one of the examiners in the first exam (23 of the 111 restorations were scheduled for replacement on all 4 examinations, with 4 on 3 occasions and only 1 on 1 occasion); again this is highlighted in Figure 4.1. BC did not schedule 7 restorations for replacement at the second examination and included 1 that was not originally scheduled for replacement. RM did not schedule 5 for replacement at the second exam and included 2 that were not originally scheduled for replacement. The differences between examinations could not be attributed to anything specific. It is possible that the examiners had undergone a subconscious recalibration but if this was the case then it is difficult to explain why less and not more restorations were scheduled for replacement on the second examination. The drop could be attributed to nothing more than co-incidence and no statistical determination or reason for the change was sought although it is possible that the research itself had honed the diagnostic intentions of the gold standard assessors. It was also possible that by setting the gold standard the physical act of self questioning, the setting of a standard and involvement in selecting suitable images for training material honed the diagnostic ability of the gold standard assessors and this potential focussing of diagnostic intention may be a "phenomenon" that requires careful investigation. It not being unrealistic to assume that the closer you look for something the more likely you are to find it; in this case a restoration requiring replacement.

As seen in Table 4.1.3 the level of agreement to the gold standard (0.77 to 0.98) suggests that an appropriate method had been employed. The inter-examiner agreement for the gold standard was in the substantial to almost perfect

categories of Landis and Koch (1997) when evaluated; 0.82 for BC and 0.77 for RM being returned when the first evaluations were compared to the gold standard determined. This correlates well with some of the standards quoted in the literature (Poulson, 1987; Ryge and Snyder, 1973, Cvar and Ryge, 1971) and exceeds that achieved in other investigations (Edwards et al, 1982; Goepferd and Kerber, 1980; Mjör and Haugen, 1976, Abou-Rass, 1973; Natkin and Guild, 1967). The overall diagnostic accuracy of the gold standard assessors as illustrated by the intra-examiner kappa statistic (0.77 for BC and 0.79 for RM) was higher than that achieved in other studies investigating replacement decisions on amalgam restorations which has reported scores ranging from 0.05 to 0.65 (Ermis and Aydin, 2004; Poorterman *et al*, 1999; Maupomé, 1998; Robertello and Pink, 1997). It needs to be remembered that it was the determination of the gold standard that was a most pertinent factor in this study and that this was determined, in the end, by consensus. It is also possible that the overall simplicity of the research i.e. limiting the evaluations to relatively simple restorations was the reason for the high level of examiner agreement between the gold standard assessors. The substantial agreement within the gold standard assessors may also have been due to a familiarity and comfortableness with the USPHS criteria and reflected in their clinical experience and education level which has been shown to have positive effects on examiner agreement (Garbuz *et al*, 2002; Robertello and Pink, 1997). The effects of applying the USPHS criteria to more complex restorations and by less experienced operators are an area that should be explored.

As shown in Tables 4.1. 3 and 4.1.4 this part of the study demonstrated that it was possible to get good agreement between two well-trained and experienced dental academics with agreements being recorded in all the statistical tests; Dices

coincidence index values (0.82 – 0.99), sensitivity (0.89 - 0.96), specificity (0.91 – 1.0), positive predictive (0.75 – 1.0) and negative predictive values (0.96 -0.98) were consistently high. This finding of consistency (and the personal acceptability of the training methods used) heavily influenced the way in which the other assessors, used in the research, were later trained. The difference in sensitivity and specificity, which is similarly noted in the positive and negative predictive values, suggests that it was easier for the assessors to determine when re-restoration was not required than when it was warranted. This is a similar finding to that reported in radiographic examinations for caries (Pretty and Maupomé, 2004c). This observation was not a surprise as the restorations used in this part of the study, although classified as simple, represented a diagnostic challenge i.e. the decision to replace was not always clear cut. This apparent easiness to decide on soundness of a restoration was also highlighted by a number of observations made by the untrained assessors in the simulated clinical phase and this is discussed later.

5.1c The recruitment and selection of dentists to take part in the study

To be acceptable, research which sets out to validate a clinical tool should be trialled in a sample that is representative of the practising population to which the research is directly applicable. Ideally, in this case, this research should be directed towards primary dental care practitioners but for practical reasons had to use clinicians who worked at the University Dental Hospital of Cardiff. This research involved part-time general dental practitioners and full-time dental academics and this was felt reasonable for the following reasons; the research was a feasibility study into the use of USPHS as a diagnostic tool, the group

allowed for flexibility (yet control) and it helped simplify Trust and Ethics Panel approvals as all volunteers already had employment contracts. The use of volunteers in any research project has a potential to introduce bias as the sample itself becomes naturally self selective. The limitations present in the group of practitioners recruited, its potential bias and how it affects external validity has to be acknowledged and the conclusions drawn suitably curtailed until similar research with other groups of practitioners has been completed. Unfortunately there is way of recruiting people to take part in any research without their co-operation and this limitation has to be accepted. An attempt to reduce bias was made by the random selection of a test and group control from the original group of volunteers; as opposed to merely accepting the first suitably interested parties. However, it has been reported in the USA that there is not much difference between dental practitioners working in dental hospitals and those working in the general dental practice environment when it comes to restorative treatment recommendations (Bader and Shugars, 1995). Figure 4.2 illustrates that the volunteer dentists were experienced (approximately 22 years of experience on average). This compares favourably with an estimated numbers of years of experience of UK registered dental practitioners as of 1st Jan 2009 (approximately 21 years, Appendix 5). It is recognised that the groups in this study may not be representative of the average general dental practitioner or average clinical academic but there is really no reason to assume that they may have fared any better or any worse than similar groups of assessors recruited from elsewhere. It is also possible that dentists who choose to work part-time in hospitals are most likely to be experienced practitioners or who exhibit preferences different to those

who choose not to work in hospitals. Only further research would confirm the aforementioned concerns.

Unlike some previous studies validating research tools (Ermis and Aydin, 2004; Forgie *et al*, 2002; Deery *et al*, 2001; Rytömaa *et al*, 1983; Poulsen *et al*, 1980) demographic data and other details were collected in the present study. This allowed comparisons to be made between the untrained and trained groups and helped confirm the heterogeneous nature of the dentists who took part. Again, only further research would be able to identify whether or not there are measurable differences in decision making amongst different generic groups e.g. private practitioners, location of practice, speciality, use of magnification, etc. Although the effect that dentist factors has on the outcome of direct restorations placed within the general dental services has been evaluated with age, country of qualification and employment status being shown to be influential (Lucarotti *et al*, 2005). The influence of factors such as magnification have as yet to be determined in general dental practice although one study carried out on dental students suggests that magnification does not unduly influence the quality of amalgam restorations produced (Donaldson *et al*, 1998). The effect that modifiers can have on the results achieved in clinical research has been previously covered (Section 2.2.3)

Sixteen practitioners were used in this study in order to facilitate data handling and provide a meaningful sample to test a new hypothesis through data gathering and statistical analysis. Sixteen were chosen as it facilitated even and manageable splits in the research protocol and allowed for the reasonable evaluation of a new technique in a sample which was necessarily constrained by resource, manageability, practicality and time. The principal researcher had also

been involved in other similar research protocols which had shown the workability and manageability of such a group size (McAndrew and Longbottom, 1993). The only limitation placed on the volunteer dentists was that they should never have participated in any form of restoration evaluation programme previously. This step was taken in order to try and prevent bias towards restoration evaluation and allow meaningful comparisons between the trained and untrained groups later in the research.

Analysis of the results (Section 4.2, Tables 4.2.1 and Tables 4.3.1 to 4.3.4 and the associated figures) showed that there was heterogeneity within the practitioners recruited in that there were no major discernable differences with respect to gender (8:8), experience (7:9), magnification (9:7) and primary place of work (10:6). This heterogeneity was further confirmed when the randomly allocated test and control groups were analysed statistically. This is discussed later.

5.2 The simulated-clinical phase

5.2a The assessment of the restorations by the untrained dentists

Before considering and discussing the results achieved by the untrained dentists it is important to highlight that on no occasion was “normal diagnostic practice” for the volunteers interfered with. Examination time was unrestricted and use of magnification neither encouraged nor discouraged; the researchers being keen to allow the decision making process for the volunteers to be as realistic and genuine as possible (within the limitations of the project). The diagnostic parameters of time and magnification use were recorded so that their influences could be determined. As with the gold standard assessors the restorations were

individually assessed having being presented in a random order at all times lest they should remember specific restorations or recall previous decisions at a later date. While each jaw was examined in turn it could be argued that a non-sequential examination procedure would have reduced the possibility for the examiners to remember specific restorations. This could be tested at a later date to see if it creates a difference.

The diagnostic challenges faced by the untrained dentists are best illustrated by considering some of the results gathered in this phase of the research (Tables 4.3.1 and 4.3.2). Of the 28 restorations scheduled for replacement by the gold standard assessors only 9 found 100% agreement for replacement with the untrained assessors and Figure 5.3 shows some of these restorations. There were 19 occasions where at least one of the untrained assessors disagreed with the gold standard for replacement (the disagreements ranged from 2 to 14). Further analysis of the “disagreement” with the gold standard assessors’ deliberations for restoration replacement showed 8 or more of the assessors to agree with the gold standard on 16 of the 19 occasions and less than 8 on only 3 occasions (Figure 5.4). Closer examination of the results did not reveal any particular pattern as to why the restorations had been scheduled for replacement by some of the assessors but not others. Of the 83 restorations not scheduled for replacement by the gold standard 100% agreement was reached for 33 of these and examples are shown in Figure 5.5. With reference to the “disagreements” 8 or more assessors disagreed with the gold standard assessors on 7 of the 50 occasions and less than 8 on 43 occasions. Figure 5.6 gives examples of restorations where more than half of the untrained assessors wanted to replace a restoration that the gold standard suggested should not be replaced.

Such findings are not new; Elderton (1976b) reporting such extremes in diagnosis over 30 years ago. His study noting that after the evaluation of 228 restorations one clinician indicated that 119 needed replacement whereas another assessor only scheduled 28. Similar variation has been reported by others (Poorterman *et al*, 1999; Bader and Shugars, 1993). The above observations, along with the findings in this study, are of paramount importance as it was felt that it needed to establish the effect that training could have in the so called “grey” areas rather than those which could be considered “black” or “white”. It has been shown that is these so called “grey” areas that cause the greatest concern (Rudolophy *et al*, 1995; Kidd and O'Hara, 1990). The apparent diagnostic difficulty experienced by the assessors in this study is consistent with other studies that have used purely visual/tactile examinations. It being shown that that the diagnosis of faulty restorations is less easy than the diagnosis of sound restorations (Bader *et al*, 2001; Merrett and Elderton, 1984; Tveit and Espelid, 1992). This may be an area for further study with efforts being targeted towards getting agreement on what is sound rather than that which is not.

5.2b The selection and training of the dentists

To prevent selection bias and allocate the assessors as randomly as possible to test and control groups selection was made by drawing numbered Scrabble™ tiles from a bag. The numbers corresponded to a numbered alphabetical list of the sixteen assessors. Scrabble™ tiles were used as they presented a uniform shape and were tactilely indistinguishable from each other. As previously detailed in Table 4.3.4 the selection process produced no discernable differences between the two groups when the results for the initial simulated clinical phase was

compared; the number of restorations scheduled for replacement ($p = 0.46$), Dice's coincidence values ($p=0.87$ and 0.39), sensitivity (0.67), specificity (0.53), positive and negative predictive values ($p=0.52$ and 0.56), Kappa statistics (0.53), washout period ($p=0.49$) and time taken for the examinations ($p=0.75$ and 0.83).

5.2c Manufacture of training material

Photographs taken at chair side by the principal investigator (RM) or by the audio visual arts department in the dental hospital were used to create a digital photographic collection and subsequent training programme relating to restoration evaluation and based on the USPHS criteria. A training booklet for use within the training programme (Appendix 2.3a to e) was created from an assessment of the photographic material by the gold standard examiners (RM and BC) with suitable representations of restoration failure being chosen. The use of photographic material in training programmes has been shown to have potential in screening programmes (Mills and Hollis, 1997). Whenever possible the clinical photographs were taken at chair side by the principal researcher. There were however occasions when this was not possible (heavy clinical commitment or failure to pre-charge the camera batteries!) realistically however there was little difference between the photographs taken at the chair side and by those taken in the AVA department. It is arguable that the advent of digital imaging and advances in ring flash usage helps eliminate some of the problems experienced when utilising conventional 35mm film where film types and conditions can influence image capture.

5.2d The training programme

Training examiners to an acceptable level of agreement is difficult (Scruggs *et al*, 1989; Weaver and Saeger, 1984; Poulsen *et al*, 1980; Houpt and Kress, 1973; Hinkelman and Long, 1973) although it has been shown, that examiner agreement within acceptable limits can also take place without training (Goepferd and Kerber, 1980; Mjör and Haugen, 1976; Abou-Rass, 1973). A formalised training and calibration programme that strived to reach 85% agreement between the dentists as suggested by Rytömaa *et al* (1983) was not felt to be applicable to this research which was primarily concerned with the effects that a basic level of training had. It has to be remembered that training and calibration are not synonymous and there are many instances where clinical studies lack evidence to show that examiners have been calibrated (Warren *et al*, 2002; Forgie *et al*, 2002; Nyvad *et al*, 1999; Ismail *et al*, 1992; Pitts and Fyffe, 1987) or indeed report results for intra- and/or inter-examiner agreement (Chesters *et al*, 2002; Deery *et al*, 2000). These comments have particular relevance to studies examining replacement decisions as it is well known that dentists have difficulty in agreeing with colleagues and peers and very often do not make the same decisions with respect to the need for dental treatment or the treatment proposed by them (Bader and Sugars, 1993; Elderton, 1986). Therefore, in comparative research involving clinical examinations, it makes sense that they are carried out under the same clinical conditions with the evaluators being trained to carry out their evaluations in a similar way; without this then discrepancy in examination is highly likely. In this study all evaluations were carried out under controlled clinical conditions with the same equipment being provided on all occasions. The training programme delivered to the principal author had produced excellent results and it was believed

that the delivery of a similar pragmatic didactic and “hands-on” training programme, with appropriate mentoring (including the use of a validation of measurement programme alongside an experienced operator) was a sensible approach. A similar approach had been successfully used in caries assessment (Deery *et al*, 1995) and amalgam restoration assessment (Robertello and Pink, 1997).

In order to allow for a suitable washout, prevent any possibility of restoration familiarity and develop the training material the training programme was delivered six months after the initial restoration evaluations. In order to provide consistency the training programme was delivered by one of the principal researchers (RM) and was timed to last no longer than one hour; this step being taken in order to reduce the possibilities of tutor and pupil fatigue. Basic understanding of the evaluation process was “checked” by the supervised assessment of a number of clinical photographs before progressing to clinical models. After the training programme, the “trainees” were given a “self-help manual” (Appendix 2a to e) illustrating the procedures involved in the assessment process and asked to review these before future assessments and asked to contact the principal author if there were any areas of confusion or misunderstanding.

It was recognised at the outset that not all “trainees” would learn at the same rate or indeed respond uniformly to the training programme that was delivered and it may be that a number of different media could have been utilised to supplement or improve the training experience e.g. computer aided learning programme, audiovisual aids, one-to-one tuition or even a web based resource e.g. <http://www.dent.umich.edu/cer/> . The appropriateness of delivering different

programmes for training in restoration assessment is an area that may warrant further investigation. To indicate whether or not a training programme, such as the above, is relevant or even practical amongst a large group of dental practitioners will require another study.

5.2e Evaluating the effects of training in the simulated clinical phase

The raw data and results from the after training during the simulated clinical phase of the project can be seen in Table 4.4.1; the corresponding results from the same assessors and the control assessors from the pre-training simulated clinical phase are also shown. As Table 4.4.1 shows the most striking finding related to the time taken by the assessors to make their decisions during this phase of the research; it was considerably longer than the time they (or indeed the control group) had taken during the pre-training simulated clinical phase and took approximately 50% longer, 59 as opposed to 39 minutes for the whole examination and 15 as opposed to 10 minutes for the kappa statistic evaluation. This increased examination time was noted despite the slight reduction in the overall number of restorations assessed (one of the models was used as a training model so 105 as opposed to 111 restoration were assessed). It appeared that the assessors in the post-training simulated clinical phase were more deliberate in their decision and thought making process in the areas examined through the use of the USPHS system. This was not surprising, as during this phase the assessors were not only expected to determine whether they would replace the restoration but grade the quality of these restorations for each of the criteria which make up the USPHS system. As, at this point in time, the assessors were relatively unfamiliar with this system; it is not surprising to see that they took longer. There were occasions

when the assessors sought clarification of the scoring system or they referred to their “self-help manual” and this obviously added to the length of the examination process. It was also felt necessary for the test (trained) assessors to verbalise their deliberations on the condition of the restorations; this was used in attempt to increase familiarity with the USPHS criteria and attempt to increase understanding of the parameters examined. This development could be likened to the changes that a student goes through as they progress from a beginners skill level through the learner phase to eventual competency. The literature failed to highlight any similar findings in clinical examinations when new tools are put into practice for the first time. However, on both occasions the increased length of time did not result in any significant differences in the number of restorations scheduled for replacement (30 compared to 34, $p = 0.4$), Dice’s Coincidence Index for agreement that restorations were sound ($p = 0.39$), Dice’s Coincidence Index for agreement that restorations needed replacing ($p = 0.12$), specificity ($p = 0.89$), positive predictive value (0.78) and negative predictive value (0.78) (Tables 4.4.5 – 4.4.11).

Analysis of the results for sensitivity (Table 4.4.8, Figures 4.4.8a-b) and the kappa statistics (Table 4.4.12 and 4.4.13) showed significant differences. The difference in sensitivity between the test (trained) group and untrained (control) group ($p = 0.021$) suggesting that the training programme had a direct and positive effect on the assessor’s ability to recognise restoration deficiencies. Notably, this difference was not mirrored in the same comparison between the untrained (test) group and trained (test) group ($p = 0.93$) and this may be caused by to the spread of the results. This finding is interesting as a number of studies suggest diagnosing faulty restorations is more difficult than diagnosing sound restorations (Bader *et al*, 2001; Tveit and Espelid, 1992; Merrett and Elderton, 1984) with

improvements in sensitivity being observed when a clinical evaluation tool is used potentially justifying its use. The apparent improvements could also be explained by the trained assessors making a conscious effort to show improvement following their training. In fact, it was noticeable that the volunteers were “hungry” for their results and indeed were keen to see how they compared to their peers. However, as none of the volunteers knew at the beginning of the research that a training programme was forthcoming it is impossible to speculate that they did not try to impress and do their best in the pre-training simulated clinical phases.

As the test group and the gold standard assessors were using the same evaluation criteria it is not surprising that the agreement between the test group and the gold standard assessors converged during this phase; the inter-examiner kappa statistic value being substantial (0.73) after training as opposed to moderate (0.57) being recorded in the pre-training simulated clinical phase. It has already been reported previously that, the assessment of restoration failure is difficult and that dentists disagree when this type of evaluation takes place but that this can be reduced with suitable training (Natkin and Guild, 1967; Abramowitz, 1966).

It should be noted that the untrained (control) assessors did not participate, at any point, in the post-training simulated clinical phase. This was necessary as the researchers did not want to compound any of the pre-training simulated clinical phase evaluations by highlighting or even subjectively influencing the untrained assessors' beliefs in restoration replacement by repeating examinations unnecessarily. A potential weakness with this approach was that there was no way of substantiating that change in the performance of this group could have taken place naturally i.e. that simple repeating of evaluations without training could enhance performance. Although anecdotal, it is well accepted that the more often

you look for something the more likely you will find it and again, arguably, the closer and harder you look the more likely you are to question yourself or your findings. This proposition could also have held true for the trained assessors i.e. it is possible that the actual training process could have unwittingly led to them becoming over or under-prescribers in restoration replacement subsequent to their training. This does not appear to be the case as the number of restorations scheduled for replacement after training was not significantly different from those before training in the simulated clinical phase; this being 27 and 28% of the total sample sizes (27.6 ± 7.44 compared with 31.3 ± 10.36). In retrospect it may have been more prudent to include a separate group of 8 untrained assessors during this phase and indeed have included them in the final evaluations. This, however, would have increased the number of assessors needed to 24 and potentially increased examination co-ordination problems at a later date; this was, therefore resisted although it is recognised as a potential deficiency in the protocol.

Table 4.4.2 showed that there was no statistically significant difference at the $p=0.05$ level for the washout times for the pre-training and post-training simulated clinical phases (18 as opposed to 57 days) for the test (trained) assessors although this result may in part be down the extreme variation shown by this group in the pre-training simulated clinical phase. Equally, no relationship between washout time and performance was observed. The “tightness” of the post-training simulated clinical phase is clear to see not only in the timeline (Figure 2.2) but also by the narrowness of the standard deviation in this group after training; 56.9 ± 47.3 days compared with 18.1 ± 4.05 days. The “control” exerted by the principal researcher (RM) in this part of the study could go some way to explaining this observation. Not only was it easier to arrange appointments for

evaluation with a smaller number of assessors but much had been learned from the pre-training simulated clinical phase and some of the leeway reduced by taking a more direct approach with the assessors involved in this phase of the study; multiple reminders to assessors were e-mailed and a more regimented approach taken e.g. the setting, checking, re-checking and co-ordination of appointments.

The results in this section of the study suggested that training could improve consistency and reliability in restoration evaluation and that the kappa statistic can be used to successfully measure such a parameter. Overall it appeared that the research agreed with the findings of others who found training had a beneficial effect (Gröndahl, 1979; Davies and Caldell, 1963; Slack *et al*, 1958). It was also clear that the delivery of such programmes is potentially both time-consuming and difficult to organise. Further it also highlights that if a similar programme was delivered and evaluated on a larger scale that considerable effort and resource would be required. It is likely that another approach to training, its delivery and its assessment may be required. As highlighted earlier there are public concerns with respect to treatment consistency provided by dentists and perhaps the introduction of acceptable programmes of restoration evaluation and assessment, once they have been validated, should be embedded into undergraduate and postgraduate training programmes; whether or not this should through conventional or alternative means remains to be determined.

5.3 The clinical phase

5.3a Recruitment of patients and the determination of the gold standard for the clinical phase

The clinical phase of the research involved the evaluation of plastic restorations in a group of patients. All examinations were conducted under the same clinical conditions and during this phase no attempt was made to segregate or separate the trained and untrained assessors when they carried out their assessments of patients and restorations. In order to simplify recruitment and facilitate the organisation of this phase of the research (where a willingness and ability to be able to consent to multiple examinations was necessary) the patients were selected from volunteers who were employees of the Cardiff and Vale NHS Trust and who worked at the Dental Hospital in Cardiff. It is noticeable that all the volunteer patients in this phase were female. This was a reflection on the gender split amongst auxiliary staff working in the dental hospital during the time of the research. On reflection, this was a good decision as a single or even a small number of male patients may have made it easier for the assessors to identify and hence improve “recall-ability” of any decisions they had made on that patient. This potential to recall decisions was a concern of the researchers at the beginning of this phase as fewer restorations were being evaluated and real people were to be used; it was felt that it would be easier to differentiate between patients than similar plastic jaws. Clearly patients have different physical attributes to differentiate them from each other and may present with occlusions and or dentitions that can, on occasion, be easily recognised. This, concern, was in fact unfounded; only one of the assessors taking part in this phase was able to recall

their decisions made during their initial examination of the patients. I will return to this point later in the thesis.

It is arguable that such a cohort may not be truly representative of the general public and presents a limited gender and age profile (19 -54 years of age, mean 30.6) . Overall, the “active” caries experience in the volunteers was considered to be lower than the national average when compared to the Adult Dental Health Survey of 1998 (Kelly *et al*, 2000) that stated “...55% of dentate adults had one or more decayed or unsound teeth.” It is also possible that this group of patients may have an increased restorative experience due to the increased access to such services. A calculation of the DMFT and comparison with national findings would confirm this. Volunteers who were actively undergoing dental treatment, who presented with minimal restorative experience, or who presented with what were felt to be only “perfect restorations” were excluded in order to lessen the likelihood that their restorations could be easily identified. Equally teeth with excessively large restorations i.e. those involving more than three surfaces were also excluded from the assessment process to eliminate difficulties that may occur when trying to assess the contact areas of restorations and be representative of those restorations used in the simulated clinical phase.

A range of restorations made from amalgam or composite resin and of variable quality were accepted for inclusion in this phase (Table 4.5.1). It is noted that nearly half of the restorations involved in this part of the study were resin based (30 of the 64). This differed significantly from that used in the simulated clinical phase which was to all intents and purposes based on amalgam restorations. Like the simulated clinical phase the reflection of the material

composition in the sample was merely a reflection on the material that was presented and collected. Although constrained by what was presented by the patients the gold standard assessors tried to ensure that the full range of the USPHS criteria was being covered in the restorations evaluated during this phase. It was necessary that the sample needed to have a level of disease that could not be easily identified by the dentists used in clinical phase but also be discernable enough to require recording; if prevalence was clearly zero then problems would arise, similarly if every restoration clearly required replacement it would not be a good test. The selection process produced an 8% incidence of restoration replacement in the patient cohort. This was considerably less than the 25% suggested by Deery *et al* (1990) as workable and to that employed in the simulated clinical phase. However, it was felt paramount that the clinical examinations should be realistic of that expected in a group of dentally motivated patients (as represented by these patients). It was also felt to be high enough to be measurable without being too obvious. As in the simulated clinical phase the prevalence of replacement rates varied between the patients and ranged from 0 to 33%. The results of the examinations confirming this as only 1 of the 14 assessors managed to be accurately recall their previous decisions.

The gold standard for restoration replacement in this phase of the study was by consensus between the gold standard assessors; the excellent agreement shown for the simulated clinical phase suggesting no need to repeat the same process for the patients. It also decreased the time required to make suitable decisions and facilitated the progress of this part of the study and minimised inconvenience to the patients.

This part of the study involved 14 of the original 16 assessors (87.5%) taking part in the research; one of the assessors was unavailable due to maternity leave, the other being unavailable due to combination of domestic and working circumstances. These absences were from both untrained and trained groups and neither presented at any point with remarkable results in the simulated clinical phases.

Table 4.5.2 details the raw data for the clinical phase and a number of statistically significant differences can be seen between the trained and untrained assessors in Table 4.5.3 with the full examination time being less for the trained group (28 compared to 37 minutes, $p= 0.003$), less restorations being scheduled for replacement by the trained group (6 compared to nearly 10, $p = 0.034$) and closer agreement with the gold standard being shown by the trained group (0.85 compared to 0.79, $p= 0.002$).

As noted, training resulted in the trained (test) group scheduling fewer restorations for replacement than the untrained (control) group. This finding suggested that the use of the evaluation criteria made the trained dentists less likely to replace a restoration. This can be explained as the training programme gave them a written description of failure to follow and hence only suggesting replacement if it fitted the description. This process is also reflected in the convergence towards the gold standard as highlighted by more favourable score for inter-examiner agreement with the gold standard. This theory is further substantiated with the finding that examination times were significantly less in the trained group of assessors. The inference being that clearer thoughts processes had led to an internalisation of the process and a more competent examination process. This is noted despite the fact that during the simulated clinical phases no

pattern emerged as to whether assessors would over-prescribe or under-prescribe with respect to their initial decision making process; after training 3 assessors replaced fewer restorations and five replaced more but yet in the simulated phase still performed no better than the control group. The results here strongly suggest that training can result in significant improvement in performance and agrees with some previous research (Robertello and Pink, 1997; Poulson, 1987; Edwards *et al*, 1982; Ryge and Snyder, 1973).

There was however no significant difference in intra-examiner agreement within the two groups which was extremely high for both the untrained (0.88) and trained groups (0.92) i.e. no group was significantly more consistent in their decision making when they were compared to themselves. However, this is not believed to be of great significance. It is believed that intra-examiner agreement, while desirable, is however not necessarily the best judge of clinical acumen or reliability i.e. a practitioner can consistently agree with themselves that a restoration requires replacement or that a tooth requires restoration but they can also be consistently wrong. It is believed that the agreement with the “gold standard” is a better marker with respect to clinical validity in decision making; in this research the trained (test) assessors did this; 0.85 compared to 0.79. The increased agreement of the trained assessors with the gold standard examiners for the restorations’ being examined is, therefore felt to be, a most valuable finding. Just as the results in the post-training simulated clinical phase showed it appears that training improves consistency in examination with respect to that determined to a gold standard. The apparent success of the training programme, although delivered in a relatively brief manner, may be explained by the fact that it contained a number of elements believed to be important in training programmes

(Weaver and Saeger, 1984): an active practical session rating samples, low levels of pre-training examiner agreement and clearly defined, well worded assessment criteria.

Although not statistically different, it is noted that on all occasions with the exception of the values calculated for positive predictive value (0.59 for both test and control groups) that there was a more favourable and higher value returned between the test and control groups for the statistical parameters evaluated. From an observational viewpoint there was a consistent trend for training to improve performance in the trained (test) group; Dices coincidence index for sound restorations (0.93 compared to 0.87), Dices coincidence index for replacement (0.37 compared to 0.35), sensitivity (0.6 compared to 0.56), specificity (0.94 compared to 0.91), negative predictive value (0.94 and 0.76). Once again, as in the simulated clinical phase the more favourable results shown for specificity, negative predictive value and the Dices coincidence index for agreement that a restoration was sound suggests that it was easier for all examiners to decide on the sound integrity of a restoration rather than one that required replacement confirming the findings of earlier research by Bader *et al* (2001), Merrett and Elderton (1984) and Tveit and Espelid (1992).

The anticipated difficulties in conducting this phase of the study were unfounded; undoubtedly the compliance of both patients and assessors were significant in the studies execution which, in hind sight was testimony to the willingness of all volunteers to participate in the research and much had been learnt in the simulated clinical phases. However the translation of similar methodology to a larger clinical study may not be so smooth unless stringent efforts are made to anticipate the difficulties of co-ordinating patient and examiner

attendances, more trained assessors are used to train un-trained volunteers or a programme of training is instigated. Indeed the possibility of instituting a national training programme in restoration evaluation at undergraduate or postgraduate level is potentially a goal for the researchers involved in this project. There would also need to be a determination made on the frequency, need and value of refresher programmes.

5.3b Confirmation of gold standard and determination of no harm

On completion of the clinical phase of the study, all patients and their associated restorations were examined by the principal researcher (RM). This served two purposes; it confirmed that the gold standard had not been altered by the repeated examinations carried out and was used to inform the patients about restorations requiring replacement; when indicated an offer to replace the restoration was made, if this was declined a letter was forwarded to the patient's general dental practitioner. While this repeat examination procedure would appear to be unique in that no reference to such a practice could be found in the literature, its practice would be justifiable from the seminal research of Ekstrand *et al* (1987) which showed that probing of occlusal surfaces can produce "irreversible traumatic defects" to teeth. It could only be considered good practice to re-assure both volunteers and participants in clinical trials, like this, that no physical harm would result from taking part in the research. It could arguably also be considered as ethically prudent measure.

5.4 An evaluation of procedure within the study by the participants

For a new clinical procedure to be accepted it needs to be safe, effective and advantageous. Additionally, it should be easy to integrate into practice and be acceptable to patients and the user. While this research showed potential in the use of USPHS, the views of the study participants were also required. A printed questionnaire was used in this project due to the relative ease of distribution, collection and subsequent analysis and interpretation. It is accepted that even when anonymity is assured there is always a degree of unreliability in drawing conclusions from survey type analyses and particularly with small surveys like this one because respondents often answer in a way that they feel they should (Munn and Drever, 1996). Additionally, an eagerness to please and provide the author with information he perhaps wanted to hear (all assessors being well known to the principal researcher and they themselves knowing the research was related to a PhD) cannot be overlooked. It is acknowledged that a focus group discussion could have determined what the volunteers thought of the research and its conduct and as a whole. Focus group research methodology has been used successfully in dentistry (Bennett *et al*, 1999; Evans *et al*, 1999; Lam *et al*, 1999; Ashford, 1998; Hastings *et al*, 1998). In this research, the group sizes of 8 are perceived as an ideal number for conducting such qualitative research (Chestnutt and Robson, 2001). With hindsight a properly conducted and transcribed focus group discussion might have explored and developed the participants' thoughts and feelings with respect to procedure and USPHS in a way that may not be forthcoming from the use of a questionnaire even when there was scope and space for free comments to be made.

As seen in Table 4.6.1 the survey revealed a number of interesting findings. Amongst the trained (test) assessors, at least three quarters rated the evaluation criteria to be easy or very easy to use. They felt that the evaluation criteria were useful and had a place in everyday clinical practice. These trained assessors also felt that their own consistency in restoration evaluation had improved. In addition to this, five of the eight taking part in the research felt they had changed their every day decision-making practice when it came to restoration replacement. Such findings suggest an acceptability of, and ease in, the applicability of the USPHS criteria in the short term. Clearly only a follow up interview could determine if they were still using the criteria today. It would also be interesting to have the study participants re-evaluate the models to determine how well they still apply the criteria. Interestingly, the results from the questionnaire confirmed the author's belief that the criteria were easy to apply even though some such as marginal discoloration and anatomical form appeared to cause the assessors some difficulty; it being noted in the simulated clinical study that the trainee's deliberations in these fields were made with less certainty.

While the questionnaire was not appropriate for the untrained (control) group their views were sought to see if they felt that their involvement in the project had altered the way they currently viewed restoration replacement. In order to ascertain this, the untrained (control) group was interviewed individually. In this instance, half (four of the eight participants) said that simply taking part in the study had indeed affected how they viewed restoration replacement. This finding has particular importance with respect to not allowing the untrained (control) group to participate during the post-training simulated clinical phase as the inclusion of an untrained (control) group who had already participated in the

simulated clinical phase may have inadvertently skewed the results. The use of the same results from the untrained (control) group from the pre-training simulated clinical phase could be interpreted as a flaw in the research, and with hindsight a separate control group could have been used here. The fact that half of the original group felt their restoration evaluation technique had been altered just by taking part vindicating the decision to exclude the control group from the post-training simulated clinical phase. It also argued that had the control group been directly involved in the post-training simulated clinical phase that this potential “Hawthorne²⁸ hangover” could have jeopardised the results achieved for the clinical phase. It was also noted that when interviewed the untrained (control) group that a number had felt aggrieved and disadvantaged for not being selected to receive training; three suggesting that they would like to undertake the training programme offered.

In addition to the questionnaire that evaluated how the assessors felt affected by the study, a number of free comments were generated that raised some important points (Table 4.6.2);

- the fact that training in restoration replacement or indeed calibration of dentists is recognised as a rare thing (unless you are participating in a trial). This observation is not unique to this research (Mjör, 2004).
- the willingness for people to take part in calibration perhaps should not be underestimated and ways in which such calibration programmes could be delivered needs to be determined. This may have been a direct result of using practitioners who were based at the hospital and colleagues who”

²⁸ Hawthorne effect: an increase in worker productivity produced by the psychological stimulus of being singled out and made to feel important. <http://www.nwlink.com/~donclark/hrd/history/hawthorne.html> 12.16am 9/8/08

were eager to participate and help out – despite no financial incentive being promised.

- that even a relatively simple decision such as “is caries present” or “is it not” can created difficulty and uncertainty even after training. This point reinforcing the view covered in the literature review that cares diagnosis is difficult
- that different operators will take on board new tools with differing degrees of enthusiasm.

5.5 General discussion point: Utilising USPHS as a clinical tool

The impact that the USPHS guidelines have had in dental research cannot be underestimated and it has been suggested that “Few if any methodological studies...have been cited more often and had greater scientific impact...” (Bayne and Schmalz, 2005). As an assessment tool USPHS compares favourably (if not better) to simpler evaluation systems such as that used by Lotzkar *et al* (1971) that looked at four areas (adaptation, contour, contact and occlusion) and better than more complicated evaluation tools such as that proposed by Hammons and Jemison (1967) that evaluates ten areas; anatomical carving, marginal ridge relation, contact, contour, marginal integrity; condensation; occlusion; tissue integrity; postoperative lavage, and surface smoothness which then had to be scored as excellent, acceptable or unacceptable. However, following this research, the fundamental question on how best to use the evaluation criteria in the clinical environment still needs to be established. The literature review in this thesis and Table 5.1 shows that USPHS has considerable support as a useful clinical tool and confirm its use as an indicator for good clinical research

(Chadwick et al, 2001) but it is still not a perfect tool. It lacks sensitivity when comparing similar materials (Hickel *et al*, 2006), the reliability and impact that individual examination parameters have on the overall decision to replace or leave a restoration has yet to be established e.g. caries, it is unknown as to what extent the progression of restoration defects progresses to failure (Bader and Shugars, 1996) and no method has materialised in order to pool results from different studies (Bayne and Schmalz, 2005). There is no work outside of a clinical trial environment to indicate how well USPHS performs over the life of a restoration in routine practice or how well it can be combined with other diagnostic techniques e.g. radiographs

Despite the above, this research has shown that a short training programme can shorten examination times and ensure convergence towards a gold standard for projected restoration replacement and affect the number of restorations projected for replacement. This finding is consistent with other research that suggests training is beneficial and improves examiner reliability (Robertello and Pink, 1997; Edwards *et al*, 1982; Goepford and Kerber, 1980, Mjör and Haugen, 1976, Abou-Rass, 1973; Natkin and Guild, 1967). Many of the points made above have also been aired in a review published in November of 2005 by Baynes and Schmalz who also highlight that a number of the fundamental practicalities of USPHS have been omitted, or chosen not to be utilised by certain researchers, over the years e.g. calibration. USPHS could be used as a tool to train dentists in restoration assessment and that in the absence of real, identifiable, recordable and justifiable reasons for replacement (e.g. pain) a restoration should not be replaced. The adjunctive use of radiographs or other diagnostic aids with USPHS needs to be considered where there is diagnostic

uncertainty. While Poorterman *et al* (1999) and Hintze and Wenzel (1994) believe that radiographs in the assessment of restorations have a limited clinical benefit in populations with a low caries experience there is research to show the contrary (Hopcraft and Morgan, 2005). It has also been shown that radiographs can have a detrimental effect in caries diagnosis and lead to over-treatment of carious lesions in the inexperienced (Maupomé and Sheiham, 1997); similar arguments could also be put forward for electronic caries diagnosis. It also needs to be remembered that the vast majority of dental restorations cannot be identified as clinically excellent or defective but lie somewhere between these two parameters and to what extent adding further stages in restoration assessment would have remains to be discovered.

It is believed that this research despite its recognised limitations provides significant evidence that the use of USPHS as a research tool in primary dental care merits consideration despite the challenges that research in general dental practice presents (Mjör *et al*, 2005).

5.6 TABLES FOR CHAPTER 5

Table 5.1 Research supporting USPHS use as a valuable clinical tool (adapted from Chadwick et al, 2001)

Year	Author(s)					
1998	de Arujo <i>et al</i>	Rasmusson <i>et al</i>	Thordrup <i>et al</i>			
1997	Abdalla <i>et al</i>	Hse and Wei				
1996	Abdalla and Aldahainy	Holan <i>et al</i>	Matis <i>et al</i>	Navarro <i>et al</i>	Qualtrough and Wilson	Wilson <i>et al</i>
1995	Cipriano and Santos	Powell <i>et al</i>	Rasmusson and Lundin	Sjogren <i>et al</i>	Tidehag and Gunne	
1994	Wendt and Leinfelder					
1992	Granath <i>et al</i>	Roberts <i>et al</i>	Studer <i>et al</i>	Wendt and Leinfelder		
1991	Matis <i>et al</i>	Wilder <i>et al</i>	Wilson <i>et al</i>			
1990	Lundin <i>et al</i>	Stangel and Barolet	Wendt and Leinfelder			
1989	Brunson <i>et al</i>	Capel <i>et al</i>	Lundin and Koch			
1988	Cavel <i>et al</i>	Sturdevant <i>et al</i>	Wilson <i>et al</i>			
1987	Oldenburg <i>et al</i>					
1986	Davis and Mayhew	Sturdevant <i>et al</i>	van Dijken			
1985	Beere <i>et al</i>	Straffon <i>et al</i>				
1984	Beere <i>et al</i>	Straffon <i>et al</i>				
1983	Paquette <i>et al</i>	Hamilton <i>et al</i>				
1982	Davis <i>et al</i>					
1981	Rydinge <i>et al</i>					
1980	Leinfelder <i>et al</i>	Tonn <i>et al</i>				

5.7 FIGURES FOR CHAPTER 5

Figure 5.1 Further representation of Table 4.1.1e showing the agreements and disagreements between the gold standard assessors

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	RA	RS	RB	RT	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29
2	Colour match	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
3	Marginal discoloration	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4	Anterioral form	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
5	Marginal adaptation	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
6	Core adaptation	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
7	Refractive	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
8	RA	RS	RB	RT	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29
9	Colour match	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
10	Marginal discoloration	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
11	Anterioral form	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
12	Marginal adaptation	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
13	Core adaptation	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
14	Refractive	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
15	RA	RS	RB	RT	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29
16	Colour match	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
17	Marginal discoloration	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
18	Anterioral form	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
19	Marginal adaptation	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
20	Core adaptation	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
21	Refractive	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
22	RA	RS	RB	RT	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29
23	Colour match	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
24	Marginal discoloration	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
25	Anterioral form	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
26	Marginal adaptation	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
27	Core adaptation	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
28	Refractive	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
29	RA	RS	RB	RT	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24	L25	L26	L27	L28	L29
30	Colour match	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Restorations where disagreement between the gold standard assessors occurred are ringed

Blue represents restorations where agreement occurred in all fields

Green represents restorations where agreement was in five of the six fields

Purple represents restorations where agreement was in four of the six fields

Red represents restorations where agreement was in three of the six fields

Orange represents restorations where agreement was in two of the six fields

Yellow represents restorations where agreement was in one of the six fields

Figure 5.2 Restoration 20793L6O, the only restoration not to be scheduled for replacement by both the gold standard assessors on the second examination.



Figure 5.3 Examples of restorations where none of the untrained assessors disagreed with the gold standard that the restorations required replacement

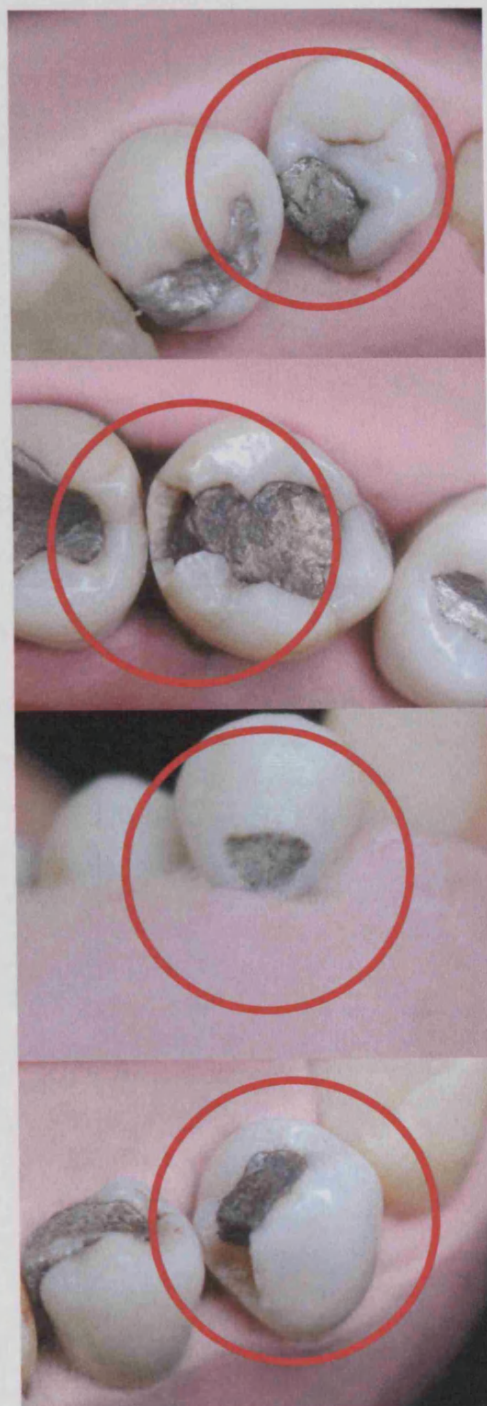


Figure 5.4 Examples of restorations where more than five of the untrained assessors disagreed with the gold standard for restoration replacement.



6 disagreed and said they would not replace the restoration



13 disagreed and said they would not replace the restoration



12 disagreed and said they would not replace the restoration

Figure 5.5 Examples of restorations where none of the untrained assessors disagreed with the gold standard that the restorations did not need to be replaced

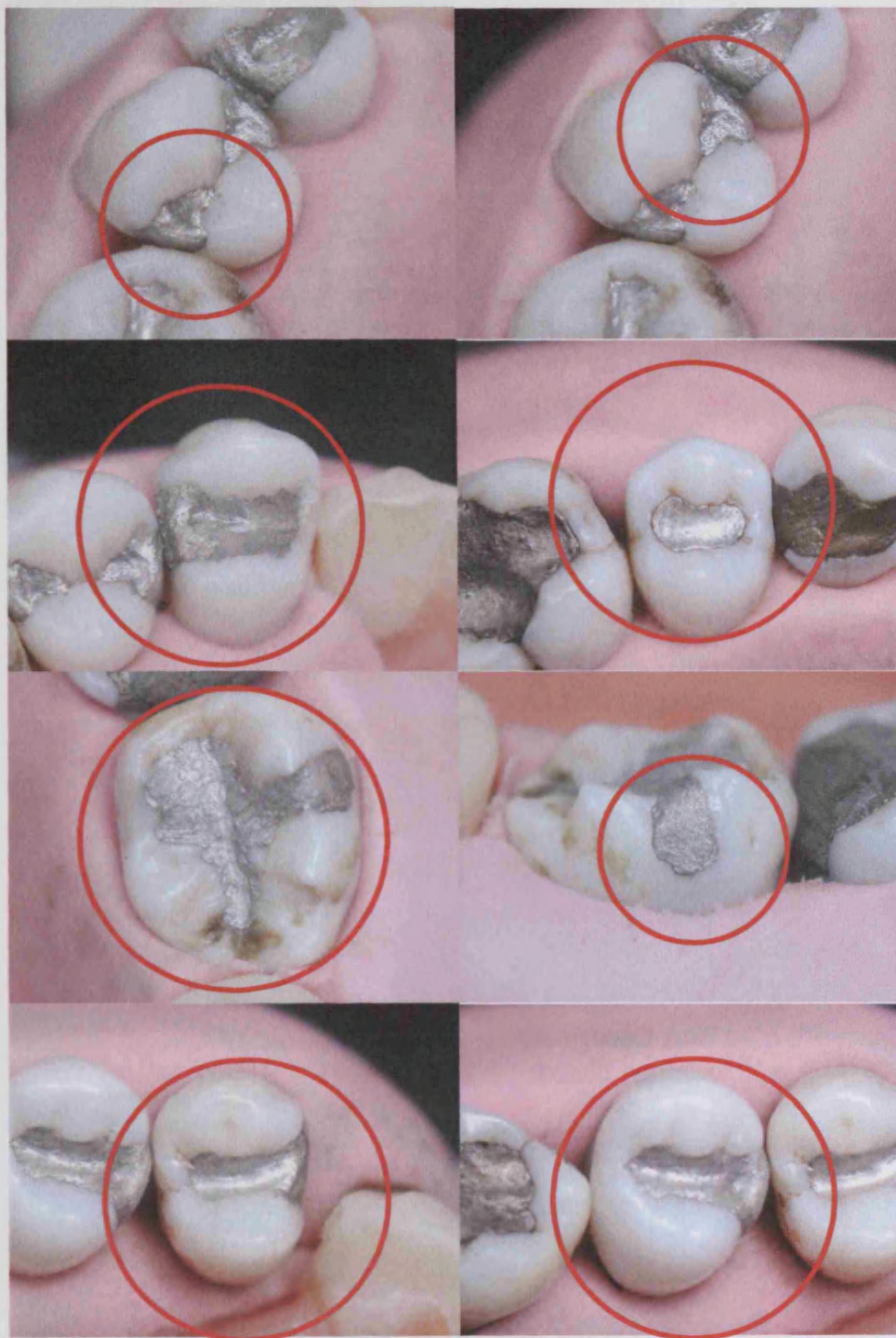


Figure 5.6 Examples of restorations where the untrained assessors' disagreed with the gold standard that the restorations did not need replaced occurred



9 disagreed and said they would replace



14 disagreed and said they would replace



11 disagreed and said they would replace



9 disagreed and said they would replace

CHAPTER 6

CONCLUSIONS AND FURTHER STUDY

6.1 CONCLUSIONS

The purpose of this study was to determine the affect that restoration evaluation training had on the decision making and restoration replacement rates amongst a group of dental practitioners. The null hypothesis stated that a restoration evaluation training programme would have no effect on any of the parameters examined. Secondary to the main purpose of the research was the determination of usefulness in using the USPHS criteria as a clinical diagnostic tool. Within the recognised and reported limitations of this study, it is believed that this research has answered the original questions with the null hypothesis being refuted and the following points noted:

- training based on the USPHS criteria can be used to deliver a restoration assessment programme,
- that a training programme increases reliability in clinical decision making (at least in the short term) and can aid the decision making process,
- restoration evaluation training leads to a quicker examination and evaluations at the chair side,
- from the limits of this study, it would appear that training leads to less restorations being scheduled for replacement and
- trained assessors agree more with the gold standard assessors.

It appears that training in restoration evaluation and assessment has an effect on restoration replacement amongst a group of experienced academics and dental practitioners and that this has significant bearing when we consider the potential costs of restoration replacement. A particular advantage of a USPHS based system being that it requires no equipment which would not normally be at the disposal of the average dental practitioner and that it does not incorporate or

evaluate aspects of restorations which they are not already familiar with. It cannot be overlooked that the training programme resulted in an increased consistency of agreement with the gold standards set and significantly decreased overall examination time in those who had undertaken the training programme.

Additionally, the relative simplicity of the USPHS criteria appeared to be well understood and that there may be potential for its use as a generic tool to help in restoration replacement decision making in dental practitioners. The delivery of a simple calibration programme in a timely and efficient manner was shown to have a statistically significant effect on the way that restorations are evaluated by those who are trained and this finding supporting the camp which suggests that training programmes positively affect clinical evaluation. It is recognised that the value that can be placed on an individual's decision making on restoration replacement based on a tool that has not been clinically validated is debatable, it cannot, however be overlooked. The old adage of "if it looks like a duck, quacks and waddles like a duck ...then it probably is a duck!" springs to mind and the undoubted experiences of many practitioners that if there is obvious caries in a tooth then it is likely to progress, if there is shadowing beyond that normally expected around a restoration that there is a real likelihood there is caries, or if dentine is exposed then dentine sensitivity is a real prospect of occurring have also to be taken into consideration when we are evaluating the value of USPHS as a diagnostically relevant tool.

It is clear from this research, the literature review and indeed much of the discussion section that there is a considerable dilemma in the longitudinal evaluation of dental restorations: we have extremes in diagnostic decision making at times and both within and between clinicians and there is as yet no agreed

methodology to best deliver training and calibration to all dental practitioners. It is also noted that despite calls for training and standardisation in restoration assessment made over the thirty or so years that we as a profession have really progressed very little. Arguably, the gradual appearance of self assessment and training modules like those produced by the University of Michigan may be one way forward but it remains to be seen what effect the availability of this medium may have in the long term. It could be that the public, their demands for better quality dentistry and value for money, the compensation culture and the increasing incidence of dental negligence claims or indeed the insistence on regulation of the profession by the statutory governing bodies, that may push practitioners to be validated in their ability to assess restorations objectively and within acceptable limits, that may take this area of interest forward. Who knows, perhaps it may be part of the practitioners educational requirements that a module of training in restorations assessment and along the same lines as radiographic exposure regulations and cardiopulmonary resuscitation training becomes part of the statutory requirements for practising dentists in the future. It is certainly hoped that this research despite its relative narrowness in execution will lead to much needed and continued evolution in this area of research and as can be seen below there are many potential areas for development. While the reasons for variability in decision making between practitioners is effectively infinite it seems sensible that if consistency in decision making is consequential to undertaking restoration evaluation training then this is a significant finding which should not be taken lightly.

6.2 FURTHER STUDY

This study identified a number of avenues for further research and some of these are highlighted:

- i. What is the best medium for delivering restoration assessment training in a cost efficient and time efficient manner and can training and calibration programmes be delivered efficiently through the medium of computer aided learning on a large scale, easily and efficiently?
- ii. What is the result of delivering the training programme in the medium and long term and how regularly should calibration programmes be re-visited for their potential benefits to remain?
- iii. Can the USPHS criteria be used, evaluated, adapted and validated for use in restoration replacement in other areas of restorative dentistry e.g. crowns and fixed prosthodontics?
- iv. Can the USPHS criteria ever be realistically validated as indicators for restoration replacement over a restorations lifetime? Does one specific criterion in USPHS influence practitioners findings more than others? Do, for example, practitioners place more relevance to the caries assessment criteria than anatomical form and would similar results be experienced to this research if it was applied to more extensive plastic restorations?
- v. Do operators who consistently replace more restorations than others benefit more or less from calibration programmes in restoration replacement?

What affect does patient history have on restoration replacement? Do chair-side magnification and the use of magnification loupes have a significant effect on restoration replacement rates? What would the overall

effect of restoration evaluation training have on the overall cost of restoration replacement?

- vi. What is the minimum period of washout between examination periods for studies like this? What influence do sample sizes have and how easy is it to recall specific findings relating to particular restorations?

CHAPTER 7

BIBLIOGRAPHY

- Abbas F, Hart AA, Oosting J *et al* (1982) Effect of training and probing force on the reproducibility of pocket depth measurements. *Journal of Periodontal Research* **17**: 226-34
- Abdalla AI and Alhadainy HA (1996) 2-year clinical evaluation of Class I posterior composites. *American Journal of Dentistry* **9**: 150-2
- Abdalla AI, Alhadainy HA and García -Godoy F (1997) Clinical evaluation of glass ionomers and compomers in Class V carious lesions. *American Journal of Dentistry* **10**: 18-20.
- Abegbembo AO and Watson PA (2005) Removal, replacement and placement of amalgam restorations by Ontario dentists in 2002. *Journal of the Canadian Dental Association* **71**: 556-65
- Abou-Rass M (1973) A clinical evaluation instrument in endodontics. *Journal of Dental Education* **37**: (Sept) 22-36
- Abramowitz J (1966) Expanded functions for dental assistants: a preliminary study. *Journal of the American Dental Association* **72**: 386-91
- Abramowitz J and Mecklenberg RE (1972) Quality of care in dental practice; the approach of the Indian Health Service. *Journal of Public Health Dentistry* **32**: 90-9
- Akerboom HB, Advokaat JG, Van Amerongen WE *et al* (1993) Long-term evaluation and reresoration of amalgam restorations. *Community Dentistry and Oral Epidemiology* **21**: 45-8
- Alhadainy HA and Abdalla AI (1996) 2-year clinical evaluation of dentin bonding systems. *American Journal of Dentistry* **9**: 77-9

- Alhadainy HA, Abdalla AI, Abegbembo AO *et al* (2005) 2-year clinical evaluation of dentin bonding systems Removal, replacement and placement of amalgam restorations by Ontario dentists in 2002. *Journal of the Canadian Dental Association* **71**: 556-65
- Allan DN (1969) The durability of conservative restorations. *British Dental Journal* **26**: 172-7
- Allan DN (1977) A longitudinal study of dental restorations. *British Dental Journal* **143**: 87-89
- Allander L, Birkead D and Bratthall D (1990) Reasons for the replacement of class II amalgam restorations in private practice. *Swedish Dental Journal* **14**: 179-84
- AlNegrish AS (2001) Reasons for placement and replacement of amalgam restorations in Jordan. *International Dental Journal* **51**: 109-15
- Ameye C, Lambrechts P and Vanherle G (1981) Conventional and microfilled composite resins. Part I: Color stability and marginal adaptation. *Journal of Prosthetic Dentistry* **46**: 623-30
- Andersson-Wenckert IE, Folkesson UH and van Dijken JWV (1997) Durability of a polyacid-modified composite resin (compomer) in primary molars. *Acta Odontologica Scandinavica* **55**: 255-60
- Anusavice JA (1988) Conference report: Criteria for placement and replacement of dental restorations. *Journal of Dental Research* **67**: 795-6
- Anusavice KJ (1989) Quality evaluation of dental restorations criteria for placement and replacement. Chicago, Quintessence Publ Co.
- Anusavice KJ (1995) Treatment regimens in preventive and restorative dentistry. *Journal of the American Dental Association* **126**: 727-43

- Ashley PF, Blinkhorn AS and Davies RM (1998) Occlusal caries diagnosis: an in vitro histological validation of the electronic caries monitor (ECM) and other methods. *Journal of Dentistry* **26**: 83-8
- Bader J and Kaplan A (1983) Treatment distributions in general dental practice. *Journal of Dental Education* **47**: 142-8.
- Bader J, Shugars D and McClure F Comparison of restorative treatment recommendations based on patients and patient simulations. *Operative Dentistry* **19**: 20-5
- Bader JD (1992) The emergence of appropriateness-of-care issues (guest editorial). *Journal of Dental Research* **71**: 502-4
- Bader JD and Brown JP (1993) Dilemmas in caries diagnosis. *Journal of the American Dental Association* **124**: (June) 48-50.
- Bader JD and Shugars DA (1992) Understanding dentists' restorative treatment decisions. *Journal of Public Health Dentistry* **52**: 102-11
- Bader JD and Shugars DA (1993) Agreement among dentists' recommendations for dental treatment. *Journal of Dental Research* **72**: 891-6
- Bader JD and Shugars DA (1995) Variation in dentists' clinical decisions. *Journal of Public Health Dentistry* **55**: 181-8
- Bader JD and Shugars DA (1996) Using crowns to prevent tooth fracture. *Community Dentistry and Oral Epidemiology* **24**: 47-51
- Bader JD, Shugars DA and Bonito AJ (2001) Systematic reviews of selected dental caries diagnostic and management methods. *Journal of Dental Education* **65**: 960-8

- Badersten A, Nilvéus R and Egelberg J (1984) Reproducibility of probing attachment level measurements. *Journal of Clinical Periodontology* **11**: 475-85
- Bailit H and Clive J (1981) The development of dental practice profiles. *Medical Care* **18**:30-46
- Bailit H, Chiriboga D, Grasso J *et al* (1979) A new intermediate dental outcome measure. *Medical Care* **17**: 780-6
- Barnes DM, Blank LW, Thompson VP *et al* (1991) A 5-year and 8-year clinical evaluation of a posterior composite resin. *Quintessence International* **22**: 143-51
- Barnes GP, Carter HG, and Hall JB (1994) Causative factors resulting in the placement of dental restorations: a survey of 8891 restorations. *Military Medicine* **138**: 736-47
- Barr-Agholme M, Odén A, Dahllöf G *et al* (1991) A two-year clinical study of light-cured composite and amalgam restorations in primary molars. *Dental Materials* **7**: 230-3
- Bates JF and Douglas WH (1980) A Two-year Field Trial of a Disperse Phase Alloy. *British Dental Journal* **149**: 133-6
- Bayne SC and Schmalz G (2005) Reprinting the classic article on USPHS evaluation methods for measuring the clinical research performance of restorative materials. *Journal of Clinical Oral Investigations* **9**: 209-14
- Beere JD, Cautley AJ and Stokes AN (1984) Composite-resin restorations of the incisal edges of anterior teeth: an assessment after 8-9 years. *New Zealand Dental Journal* **80**: 72-4

- Belcher MA and Stewart GP (1997) Two year clinical evaluation of an amalgam adhesive. *Journal of the American Dental Association* **128**: 309-14
- Bentley C and Drake C (1986) Longevity of restorations in a dental school clinic. *Journal of Dental Education* **50**: 594-600
- Berggren H and Welander B (1960) The unreliability of caries recording methods. *Acta Odontologica Scandinavica* **18**: 409-20
- Berry TG, Nicholson J *et al* (1994) Almost two centuries with amalgam: where are we today? *Journal of the American Dental Association* **125**: 392-9
- Berry TG, Osborne JS, and Summitt JB (1995) Clinical evaluation of high-copper amalgams. *American Journal of Dentistry* **8**: 122-4
- Berset GP, Eriksen HM, Bjertness E *et al* (1996) Caries experience of 35-year-old Oslo residents and changes over a 20-year period. *Community Dent Oral Epidemiol* **13**: 238-44
- Billington RW, Williams JA and Pearson GJ (1990) Variation in powder/liquid ratio of a restorative glass-ionomer cement used in dental practice. *British Dental Journal* **169**: 164-7
- Bjertness E and Sønju T (1990) Survival analysis of amalgam restorations in long-term recall patients. *Acta Odontologica Scandinavica* **48**: 93-7
- Blunck U and Roulet JF (1999) Marginal adaptation of compomer Class V restorations in vitro. *Journal of Adhesive Dentistry* **1**: 143-51
- Boksman L, Jordan RE, Suzuki M *et al* (1986) A visible light-cured posterior composite resin: results of a 3-year clinical evaluation. *Journal of the American Dental Association* **112**: 627-31

- Boyd M. Amalgam replacement: are decisions based on fact or tradition?
In: Anusavice KJ, ed. *Quality evaluation of dental restorations*. Chicago, Quintessence, 1989; 73-82
- Boyd MA and Richardson AS (1985) Frequency of amalgam replacement in general dental practice. *Journal of the Canadian Dental Association* **51**: 763-6
- Braga SR, Vasconcelos BT, Macedo NR *et al* (2007) Reasons for placement and replacement of direct restorative materials in Brazil. *Quintessence International* **38**: E189-94
- Braun RJ and Marcus M (1985) Comparing treatment decisions for elderly and young dental patients. *Gerodontology* **1**: 138-42
- Brennan DS and Spencer AJ (2003) Restorative service patterns in Australia: amalgam, composite and glass ionomer restorations. *International Dental Journal* **53**: 455-63
- Brunson WD, Bayne SC, Sturdevant JR *et al* (1989) Three-year clinical evaluation of a self-cured posterior composite resin. *Dental Materials* **5**: 127-32
- Bryant RW and Hodge KV (1994) A clinical evaluation of posterior composite resin restorations. *Australian Dental Journal* **39**: 77-81
- Burke FJT (2006) Dentists' influence on restoration longevity. *Dental Update* **3**: 69
- Burke FJT, Cheung SW, Mjör IA *et al* (1999) Restoration longevity and analysis of reasons for the placement and replacement of restorations provided by vocational dental practitioners and their trainers in the United Kingdom. *Quintessence International* **30**: 234-42

- Burke FJT, Wilson NHF, Cheung SW *et al* (2001) Influence of patient factors on age of restorations at failure and reasons for their placement and replacement. *Journal of Dentistry* **29**: 317-24
- Byar DP (1977) Sound advice for conducting clinical trials. *New England Journal of Medicine* **297**: 553-4
- Cabot L (1990) Variation in complete denture assessment. *Journal of Dentistry* **18**: 98-101
- Capel CPE, Gomes MW Jr and Ferreira SJF (1989) Low-silver amalgam restorations: a two-year clinical evaluation. *Dental Materials* **5**: 277-80
- Cardoso PEC, Miranda WG Jr. and Santos JFF (1989) Low-silver amalgam restorations: a two-year clinical evaluation. *Dental Materials* **5**: 277-80
- Cavel WT, Kelsey WP, Barkmeier WW *et al* (1988) A pilot study of the clinical evaluation of castable ceramic inlays and a dual-cure resin cement. *Quintessence International* **19**: 257-62
- Chadwick BI, Dummer PM, Dunstons FD *et al* (1999) What type of filling? Best practice in dental restorations. *Quality Health Care* **8**: 202-7
- Chadwick BL, Dummer PMH, Dunstan F *et al* (2001) The longevity of dental restorations. A systematic review. NHS Centre for Reviews and Dissemination, University of York, England.
- Charbeneau GT (1975) Rating scales for the clinical evaluation of quality of performance in restorative dentistry In *An instructional information exchange for dentistry in the United States* **6** 153-178 Mackenzie RS and Harrop TJ, eds Washington: US Govt Printing Office (DHEW Pub No (HRA) 75-8)

- Chesters RK, Pitts NB, Matuliene Get *al* (2002) An abbreviated caries clinical trial design validated over 24 months. *Journal of Dental Research* **81**: 637-40
- Christensen RP and Christensen GJ (1982) In vivo comparison of a microfilled and a composite resin: a three-year report. *Journal of Prosthetic Dentistry* **48**: 657-63
- Cipriano TM and Santos JFF (1995) Clinical behaviour of repaired amalgam restorations: a two-year study. *Journal of Prosthetic Dentistry* **73**: 8-11
- Clarkson JE, Worthington HV and Davies RM (2000) Restorative treatment provided over five years for adults regularly attending a general dental practice. *Journal of Dentistry* **28**: 233-39
- Cleaton-Jones P, Hargreaves JA, Fatti LP *et al* (1989) Dental caries diagnosis calibration for clinical field surveys. *Caries Research* **23**: 195-9
- Cohen J A (1960) A coefficient of agreement for nominal scales. *Educational Psychological Measurement* **20**: 37-46
- Collins CJ, Bryant RW and Hodge KLV (1998) A clinical evaluation of posterior composite resin restorations. *Journal of Dentistry* **26**: 311-17
- Corpron RE, Straffon LH, Dennison JB *et al* (1982) A clinical evaluation of polishing amalgams immediately after insertion: 18 month results. *Paediatric Dentistry* **4**: 98-105
- Corpron RE, Straffon LH, Dennison JB *et al* (1983) A clinical evaluation of polishing amalgams immediately after insertion: 36 month results. *Pediatric Dentistry* **5**: 126-30
- Crabb HSM (1981) The survival of dental restorations in a teaching hospital. *British Dental Journal* **150**: 315-22

- Crawford WH (1938) Amalgam failures. *New York Journal of Dentistry* **8**: 256-8
- Crombie IK (2005) The pocket guide to critical appraisal. *British Medical Journal Publishing Group* 1996, BMA House, Tavistock Square, London WC19JR.
- Cunningham J, Mair LH, Foster MA *et al* (1990) Clinical evaluation of three posterior composite and two amalgam restorative materials: 3-year results. *British Dental Journal* **169**: 319-23
- Cvar JF and Ryge G (1971) Criteria for the Clinical Evaluation of Dental Restorative Materials, San Francisco: US Government Printing Office
- Dahl JE and Eriksen HM (1978) Reasons for replacement of amalgam dental restorations. *Scandinavian Journal of Dental Research* **86**: 404-7
- Darvell BW (1978) Deterioration of disperse-phase amalgam alloy. *British Dental Journal* **21**:181-4
- Davies GN and Cadell PB (1963) Four investigations to determine the reliability of caries-recording methods. *Archives of Oral Biology* **8**: 331-48
- Davis RD and Mayhew RB (1986) A clinical comparison of three anterior restorative resins at 3 years. *Journal of the American Dental Association* **112**: 659-63
- Davis RD, Mayhew RB and Voss JE (1982) A clinical comparison of three tooth-colored restorative resins. *Journal Indiana Dental Association* **61**: 9-12
- Dawson AS and Smales RJ (1988) Restoration longevity in an Australian Defence Force population. *Australian Dental Journal* **37**: 196-200

- Deery C, Care R, Chesters R *et al* (2000) Prevalence of dental caries in Latvian 11-to-15 year-old children and the enhanced diagnostic yield of temporary tooth separation, FOTI and electronic caries measurement. *Caries Research* **34**: 2-7
- Deery C, Fyffe HE, Nugent Z *et al* (1995) The effect of placing a clear pit and fissure sealant on the validity and reproducibility of occlusal caries diagnosis. *Caries Research* **29**: 377-381
- Deligeorgi V, Wilson NHF, Fouzas D *et al* (2000) Reasons for placement and replacement of restorations in student clinics in Manchester and Athens. *European Journal of Dental Education* **4**: 153-9
- Dental Estimates Board (1980) Annual report. E. a. Wales, Eastbourne. **Appendix H**: 51-8
- Dental Estimates Board. (1981) Summary Report. Edinburgh. **Appendix I**: 1-3
- Derkson GD, Richardson AS, Waldman R *et al* (1982) Clinical evaluation of composite and amalgam posterior restorations: one year results. *Journal of the Canadian Dental Association* **48**: 45-7
- de Araujo MAM, Araujo RM, Marsilio DDS (1998). A retrospective look at the esthetic resin composite and glass-ionomer Class III restorations: A 2 year clinical evaluation. *Quintessence International* **29**: 87-93
- de Vries HC, Ruiken HM, Konig KG *et al* (1990) Radiographic versus clinical diagnosis of approximal carious lesions. *Caries Research* **24**: 364-70
- Dice LR (1945) Measures of the amount of ecological association between species. *Ecology* **26**: 297-302

- Diefenbach V (1972) The role of organized dentistry in quality of dental care. Spring Work Conference. Philadelphia, Philadelphia County Dental Society.
- Dietschi D, Ciucchi B and Holz J (1989) A clinical trial of four light curing posterior composite resins: 9-month. *Quintessence International* **20**: 641-652
- Donaldson ME, Knight GW and Guenzel PJ (1998) The effect of magnification on student performance in pediatric operative dentistry. *Journal of Dental Education* **62**: 905-10
- Douglass CW and Gammon MD (1984) The epidemiology of dental caries and its impact on the operative dentistry curriculum. *Journal of Dental Education* **48**: 547-55
- Downer MC (1989) Validation of methods used in dental caries diagnosis. *International Dental Journal* **39**: 241-6
- Downer MC, Azli NA, Bedi R *et al* (1999) How long do routine dental restorations last? A systematic review. *British Dental Journal* **187**: 432-9
- Drake CW (1988) A comparison of restoration longevity in maxillary and mandibular teeth. *Journal of the American Dental Association* **116**: 651-4
- Drake CW, Marynuik GA and Bentley C (1990) Reasons for restoration replacement: differences in practice patterns. *Quintessence International* **21**: 125-30
- Duke SE (1992) Clinical studies of adhesive systems. *Operative Dentistry* (Supplement 5):103-10
- Dunn G and Everitt BS (1995) Clinical biostatistics – an introduction to evidence-based medicine. London: Edward Arnold

- Dunston KR, Milgrom P *et al* (1978) Practitioner-based evaluation criteria for dental education. *Journal of Dentistry for Children* **45**: 207-12
- Duperon DF, Neville MD and Kasloff Z (1970) Clinical evaluation of corrosion resistance of conventional alloy, spherical-particle alloy, and dispersion-phase alloy. *Journal of Prosthetic Dentistry* **25**: 650-6
- Eames WB and MacNamara JF (1976) Eight high copper amalgam alloys and six conventional alloys compared. *Operative Dentistry* **1**: 98-107
- Eames WB, Strain JD, Weitman RT *et al* (1974) Clinical comparison of composite, amalgam and silicate restorations. *Journal of the American Dental Association* **89**: 1111-7
- Edwards WA, Morse PK and Mitchell RJ (1982) A practical evaluation system for preclinical restorative dentistry. *Journal of Dental Education* **46**: 693-6
- Eidelman E, Fuks A and Chosack A (1989) A clinical, radiographic, and SEM evaluation of Class 2 composite restorations in primary teeth. *Operative Dentistry* **14**: 58-63
- Ekstrand KR, Qvist V and Thylstrup A (1987) Light microscope study of the effect of probing in occlusal surfaces. *Caries Research* **21**: 368-74
- Ekstrand KR, Ricketts DNJ and Kidd EAM (1997) Reproducibility and accuracy of three methods for assessment of demineralization depth of the occlusal surface: an in vitro examination. *Caries Research* **31**: 224-31
- Elderton R and Eddie S (1983) The changing pattern of treatment in the general dental service, 1965-1981. Part 2- restorative treatment for the future. *British Dent Journal* **144**: 421-3

- Elderton R and Nuttall V (1983) Variation amongst dentists in planning treatment. *British Dental Journal* **154**: 201-6
- Elderton RJ (1976a) The causes of failures of restorations: A literature review. *Journal of Dentistry* **4**: 257-62
- Elderton RJ (1976b) The prevalence of failure of restorations; a literature review. *Journal of Dentistry* **4**: 207-10
- Elderton RJ (1977a) An objective assessment of the quality of amalgam restorations. In, A series of monographs on the assessment of the quality of dental care :Allred H (Ed), the London Hospital Medical College p 55-74
- Elderton RJ (1977b) Assessment of the quality of restorations: a literature review. *Journal of Oral Rehabilitation* **4**: 217-26
- Elderton RJ (1979) A new look at cavity preparation. *Proceedings of the British Paedontic Society* **9**: 25-30
- Elderton RJ (1983) Longitudinal study of dental treatment in the General Dental Service in Scotland. *British Dental Journal* **155**: 91-96
- Elderton RJ (1985) Implications of recent dental health services research and the future of operative dentistry. *Journal of Public Health Dentistry* **45**: 101-5
- Elderton RJ (1993) Overtreatment with restorative dentistry: When to intervene? *International Dental Journal* **43**: 17-24
- Elderton RJ and Eddie S (1983) The changing pattern of treatment in the general dental service, 1965-1981. *British Dental Journal* **144**: 421-3
- Elderton RJ and Nuttall NM (1983) Variation among dentists in planning treatment. *British Dental Journal* **154**: 201-6

- El-Mowafy OM, Lewis DW, Benmergui *et al* (1994) Meta-analysis on long-term clinical performance of posterior composite restorations. *Journal of Dentistry* **22**: 33-43
- Ermis BR and Aydin U (2004) Examiner agreement in the replacement decision of Class I amalgam restorations. *The Journal of Contemporary Dental Practice* **5**: 1-7
- ErtenH, Uctasli MB, Akarslan ZZ *et al* (2005) The assessment of unaided visual examination, intra-oral camera and operating microscope for the detection of occlusal caries lesions. *Operative Dentistry* **30**: 190-4
- Ettinger RL (1999) Epidemiology of dental caries. A broad review. *Dental Clinics of North America* **43**: 679-94
- Espelid I and Tveit AB (1991) Diagnosis of secondary caries and crevices adjacent to amalgam. *International Dental Journal* **41**: 359-64
- Espelid I, Tveit AB, Erickson RL *et al* (1991) Radiopacity of restorations and detection of secondary caries. *Dental Materials* **7**: 114-7
- Espevik S (1977) Creep of dental amalgam and its phases. *Scandinavian Journal of Dental Research* **85**: 492-5
- Faragher B and Marguerie C (1998) Essential statistics for medical examinations. Pastest, Egerton Court, Parkgate Estate, Knutsford, Cheshire WA16 8DX
- Fenton RA and Smales RJ (1984) Immediate-polished and as-carved Tytin restorations after 12 months. *Journal of Dentistry* **12**: 165-74
- Ferreira Zandona AG *et al* (1998) Laser fluorescence detection of demineralization in artificial occlusal fissures. *Caries Research* **32**: 31-40

- Flack VF, Atchison KA, Hewlett ER *et al* (1996) Relationships between clinician variability and radiographic guidelines. *Journal of Dental Research* **75**: 775-82
- Fontana M (1995) Relation with current criteria used to replace restorations. *General Dentistry* **43**: 145-52
- Fontana M and Gonzalez-Cabezas C (2000) Secondary caries and restoration replacement an unresolved problem. *Compendium of Continuing Dental Education* **21**: 46-54
- Forgie AH, Pine CM and Pitts NB (2002) The use of magnification in a preventative approach to caries detection. *Quintessence International* **33**: 13-6
- Forgie AH, Pine CM and Pitts NB (2003) The assessment of an intra-oral video camera as an aid to Occlusal caries detection. *Quintessence International* **53**: 497-502
- Forss H and Widström E (1996) Factors influencing the selection of restorative materials in dental care in Finland. *Journal of Dentistry* **24**: 257-62
- Foster LV (1994) Validity of clinical judgements for the presence of secondary caries associated with defective amalgam restorations. *British Dental Journal* **177**: 89-93
- Friedl KH, Hiller KA and Schmalz G (1994) Placement and replacement of amalgam restorations in Germany. *Operative Dentistry* **19**: 228-32
- Friedl KH, Hiller KA and Schmalz G (1995) Placement and replacement of composite restorations in Germany. *Operative Dentistry* **20**: 34-8

- Friedman JW (1972) A guide for the evaluation of dental care. *School of Public Health, University of California, LA. Monograph*
- Frost PM (2002) An audit on the placement and replacement of restorations in general dental practice. *Primary Dental Care* **9**:
- Fuchs VR (1978) The supply of surgeons and the demand for operations. *Journal of Human Resources* **13(suppl)**: 35-6
- Fukushima M, Setcos JC and Phillips RW (1988) Marginal fracture of posterior composite resins. *Journal of the American Dental Association* **117**: 577-83
- Fuller JL (1972) The effects of training and criterion models on interjudge reliability. *Journal of Dental Education* **36**: 19-22
- Fyffe HE, Deery C, Nugent ZJ *et al* (2000) Effect of diagnostic threshold on the validity and reliability of epidemiological caries diagnosis using the Dundee Selectable Threshold Method for caries diagnosis (DSTM). *Community Dentistry and Oral Epidemiology* **28**: 42-51
- Gaengler P, Hoyer I, Montag R *et al* (2004) Micromorphological evaluation of posterior composite restorations – a 10 year report. *Journal of Oral Rehabilitation* **31**: 991-1000
- Garbuz DS, Masri BA, Estalle J *et al* (2002) Classification systems in orthopaedics. *Journal of the American Academy of Orthopaedic Surgeons* **10**: 290-7
- Gaus CR, Cooper S *et al* (1976) Contrast in HMO and fee-for-service performance. *Social Security Bulletin* **39**: 3-14
- Gibb GH (1966) Methods for achieving improved amalgam restorations. *Journal of the South African Dental Association* **15**: 17-22

- Gibson GB, Richardson AS, Patton RE *et al* (1982) A clinical evaluation of occlusal composite and amalgam restorations: one-year and two-year results. *Journal of the American Dental Association* **104**: 335-7
- Glazer AN (1995) High-yield biostatistics. Baltimore: Williams and Wilkins
- Goepferd SJ and Kerber PE (1980) A comparison of two methods for evaluating primary class II cavity preparations. *Journal of Dental Education* **44**: 537-42
- Going RE (1972) Microleakage around dental restorations: a summarizing review. *Journal of the American Dental Association* **84**: 1349-57
- Goldberg AJ (1990) Deterioration of restorative materials and the risk for secondary caries. *Advances in Dental Research* **4**: 14-8
- Goldberg AJ, Rydinge E, Santucci EA *et al* (1984) Clinical evaluation methods for posterior composite restorations. *Journal of Dental Research* **63**: 1387-91
- Granath L, Schroder U and Sundin B (1992) Clinical evaluation of preventive and class-I composite resin restorations. *Acta Odontologica Scandinavica* **50**: 359-64
- Green LA and Dovey SM (2001) Practice-based primary care research networks: they work and they are ready for full development and support. *British Medical Journal* **322**: 567-8
- Grembowski D, Milgrom P and Fiset L (1991) Dental decision-making and variation in dentist service rates. *Social Science Medicine* **32**: 287-94
- Grogono AL, McInnes PM, Zinck JH *et al* (1990) Posterior composite and glass ionomer/composite laminate restorations: 2-year clinical results. *American Journal of Dentistry* **3**: 147-52

- Gröndahl HJ (1979) Some factors influencing observer performance in radiographic caries diagnosis. *Swedish Dental Journal* **3**: 157-72
- Grossman ES and Matejka JM (1997) Amalgam marginal quality assessment: a comparison of seven methods. *Journal of Oral Rehabilitation* **24**: 496-505
- Gruebbel AO (1950) Report on the study of dental public health services in New Zealand. *Journal of the American Dental Association*, **41**: 275-83
- Gruythuysen RJM, Kreulen KM, Tobi H *et al* (1996) 15-year evaluation of Class II amalgam restorations. *Community Dentistry Oral Epidemiology* **24**: 207-10
- Haak R, Wicht MJ, Hellmich M *et al* (2002) The validity of proximal caries detection using magnifying visual aids. *Caries Research* **36**: 3-6
- Hadley J, Holahan H *et al* (1979) Can fee-for-service reimbursement coexist with demand creation? *Inquiry* **16**: 247-58
- Haffajee AD, Socransky SS and Goodson JM (1983) Clinical parameters as predictors of destructive periodontal disease. *Journal of Clinical Periodontology* **10**: 257-65
- Hamilton JC, Moffa JP and Ellison JA *et al* (1983) Marginal fracture not a predictor of longevity for two dental amalgam alloys: a ten-year study. *Journal of Prosthetic Dentistry* **50**: 200-2
- Hammons PE, Jamison HC *et al* (1971) Quality of service provided by dental therapists in an experimental program at the University of Alabama. *Journal of the American Dental Association* **31**: 1060-3

- Hawthorne WS and Smales RJ (1997) Factors influencing long-term restoration survival in three private dental practices in Adelaide. *Australian Dental Journal* **42**: 59-63
- Hayashi M and Wilson NH (2003) Failure of posterior composites with post-operative sensitivity. *Operative Dentistry* **28**: 681-8
- Hazelkorn H (1995) A comparison of dental treatment plans under different reimbursement systems. *Journal of Public Health Policy* **6**: 223-35
- Hazelkorn H and Macek M (1994) Perception of the need for removal of impacted third molars by general dentists and oral and maxillofacial surgeons. *Journal of Oral and Maxillofacial Surgery* **52**: 681-6
- Healey HJ and Phillips RW (1949) A clinical study of amalgam failures. *Journal of Dental Research* **28**: 439-46
- Hedegård B (1955) Cold-polymerizing resins. *Acta Odontologica Scandinavica* **13**: suppl, 17. 57
- Heintze SD (2007) Systematic reviews II: The correlation between marginal quality and clinical outcome. *Journal of Adhesive Dentistry* **9**: suppl 1, 77-106
- Hembree JHJ and Andrews T (1980) Microleakage evaluation of eight composite resins. *Journal of Prosthetic Dentistry* **44**: 279-82
- Hendriks FHJ, Letzel H and Vrijhoef MMA (1985) Cost benefit analysis of direct posterior restorations. *Community Dentistry and Oral Epidemiology* **13**: 256-59
- Hendriks FHJ, Letzel H and Vrijhoef MMA (1986) Composite versus amalgam restorations. A three-year clinical evaluation. *Journal of Oral Rehabilitation* **13**: 401-11

- Hesselgren K and Thylstrup A (1982) Development in dental caries among children in 1961-79 in a Danish community with school dental service. *Community Dentistry and Oral Epidemiology* **10**: 276-81
- Hewlett ER, Atchison KA, White SC *et al* (1993) Radiographic secondary caries prevalence in teeth with clinically defective restorations. *Journal of Dental Research* **72**: 1604-08
- Hickel R and Manhart J (2001) Longevity of restorations in posterior teeth and reasons for failure. *Journal of Adhesive Dentistry* **3**: 45-64
- Hickel R, Dasch W, Janda RI *et al* (1998) New direct restorative material. *International Dental Journal* **48**: 3-16
- Hickel R, Manhart J and Garcia-Goody F (2000) Clinical results and new developments of direct posterior restorations. *American Journal of Dentistry* (Spec no): 41D-54D
- Hickel R, Roulet JF, Bayne S *et al* (2007) Recommendations for conducting controlled clinical studies of dental restorative materials. *Clinical Oral Investigations* **11**: 5-33
- Hinkelman KW and Long NK (1973) Method for decreasing subjective evaluation in preclinical restorative dentistry. *Journal of Dental Education* **37: (Sept)** 13-18
- Hintze H and Wenzel A (1994) Clinically undetected dental caries assessed by bitewing screening in children with little caries experience. *Dentomaxillofacial Radiology* **23**: 19-23
- Hintze H, Wenzel A, Danielson B *et al* (1998) reliability of visual examination, fibre-optic transillumination, and bite-wing radiography, and reproducibility of direct visual examination following tooth separation for the

identification of cavitated carious lesions in contacting approximal surfaces.

Caries Research **32**: 204-9

- Hintze H and Wenzel A (2002) Influence of the validation method on diagnostic accuracy for caries. A comparison of six digital and two conventional radiographic systems. *Dentomaxillofacial Radiology* **31**: 44-9
- Hintze H and Wenzel A (1999) A two-film versus four-film bite-wing examination for caries diagnosis in adults. *Caries research* **33**: 380-6
- Hocott J (1984) Treatment planning and management. *Journal of Dental Education* **48**: (Suppl 6) 13-6
- Holan G, Chosack A and Eidelman E (1996) Clinical evaluation of Class II combined amalgam-composite restorations in primary molars after 6 to 30 months. *Journal of Dentistry for Children* **63**: 341-5
- Hopcroft MS and Morgan MV (2005) Comparison of radiographic and clinical diagnosis of approximal and occlusal dental caries in a young adult population. *Community Dentistry and Oral Epidemiology* **33**: 212-8
- Horowitz HS, Baume LJ, Backer-Dirks O *et al* (1973) Principal requirements for controlled clinical trials of caries preventive agents and procedures. *International Dental Journal* **23**: 506-16
- Houpt MI and Kress G (1973) Accuracy of measurement of clinical performance in dentistry. *Journal of Dental Education* **37**: 34-46
- Hse KM and Wei SH (1997) Clinical evaluation of compomer in primary teeth: 1-year results. *Journal of the American Dental Association* **128**: 1088-96

- Hugoson A, Koch G, Slotte C *et al* (2000) Caries prevalence and distribution in 20-80-year-olds in Jonkoping, Sweden, in 1973, 1983, 1993. *Community Dentistry and Oral Epidemiology* **28**: 90-6
- Hunter B (1985) Survival of dental restorations in young patients. *Community Dentistry and Oral Epidemiology* **13**: 285-7
- ICDAS (2005) Rationale and evidence for the International Caries Detection and Assessment System (ICDAS II). <http://icdas.org> Retrieved July 31, 2006
- Ismail AI (2004) Visual and visuo-tactile detection of dental caries. *Journal of Dental Research* **82**: C56-66
- Ismail AI, Brodeur JM, Gagnon P *et al* (1992) Prevalence of non-cavitated carious lesions in a random sample of 7-9-year-old schoolchildren in Montreal, Quebec. *Community Dentistry and Oral Epidemiology* **20**: 250-55
- ISO (2003) Dental materials- Testing of adhesion to tooth structure. Technical specification No: 11405
- Jacobsen PH (1988) Design and analysis of clinical trials. *Journal of Dentistry* **16**: 215-8
- Jodkowska E (1985) 2 years' clinical evaluation of fillings made of Biotrey and Silicap materials. *Czasopismo Stomatologiczne* **38**: 173-77
- Johnson GH, Bales DJ, Gordon GE *et al* (1992) Clinical performance of posterior composite resin restorations. *Quintessence International* **23**: 705-11
- Jokstad A and Mjör IA (1991a) Analyses of long-term clinical behaviour of class-II amalgam restorations. *Acta Odontologica Scandinavica* **49**: 47-63

- Jokstad A and Mjör IA (1991b) Replacement reasons and service time of class-II amalgam restorations in relation to cavity design. *Acta Odontologica Scandinavica* **49**: 109-126
- Jokstad A, Bayne S, Blunck U *et al* (2001) Quality of dental restorations *International Dental Journal* **51**: 117-58
- Jokstad A, Mjör IA and Qvist V (1994) The age of restorations in situ. *Acta Odontologica Scandinavica* **52**: 234-42
- Jordan RE, Suzuki M and Davidson DF (1993) Clinical evaluation of a universal dentin bonding resin: preserving dentition through new materials. *Journal of the American Dental Association* **124**: 71-6
- Jorgensen KD and Wakumoto S (1968) Occlusal amalgam filings marginal defects and secondary caries. *Odont Tidsk* **76**: 43-54
- Kalsbeek H, Truin G-J, Poorterman JHG *et al* (2000) Trends in periodontal status and oral hygiene habits in Dutch adults between 1983 and 1995. *Community Dentistry and Oral Epidemiology* **28**: 112-8
- Kaste LM, Selwitz RH *et al* (1996) Coronal caries in the primary and permanent dentition of children and adolescents 1-17 years of age: United States, 1988-1991. *Journal of Dental Research* **75 (Spec no)**: 631-41
- Kay E and Nuttall N (1994) Relationship between dentists' treatment attitudes and restorative decisions made on the basis of simulated bitewing radiographs. *Community Dentistry and Oral Epidemiology* **22**: 71-4
- Kay E, Nuttall N and Knill-Jones R (1992) Restorative treatment thresholds and agreement treatment decision-making. *Community Dentistry and Oral Epidemiology* **20**: 265-8

- Kay E, Watts A, Paterson R *et al* (1988) Preliminary investigation into the validity of dentists' decision to restore occlusal surfaces of permanent teeth. *Community Dentistry and Oral Epidemiology* **16**: 191-4
- Kelly M, Steele J, Nuttall N *et al* (2000) Adult Dental Health Survey: Oral Health in the United Kingdom 1998. *Office for National Statistics*, the Stationery Office, London. ISBN 011 621268 3
- Kelsey WP, Franco SJ and Blinkhorn AS (1981) Caries as a cause of restoration replacement: a clinical survey. *Quintessence International* **12**: 971-4
- Kennon S (1989) Faculty alterations of patient treatment plans. *Journal of Dental Education* **53**: 587-9
- Ketley CE and Holt RD (1993) Visual and radiographic diagnosis of caries in first permanent molars and in second primary molars. *British Dental Journal* **174**: 364-70
- Kidd EAM (2001) Diagnosis of secondary caries. *Journal of Dental Education* **65**: 997-1000
- Kidd EAM and Beighton D (1996) Prediction of secondary caries around tooth-coloured restorations: A clinical and microbiological study. *Journal of Dental Research* **74**: 1942-6.
- Kidd EAM, Joyston-Bechal S and Beighton D (1995) Marginal ditching and staining as a predictor of secondary caries around amalgam restorations: a clinical and microbiological study. *Journal of Dental Research* **74**: 1206-11
- Kidd EAM, Ricketts DNJ and Pitts NB (1993) Occlusal caries diagnosis: A changing challenge for clinicians and epidemiologists. *Journal of Dentistry* **27**: 409-16

- Kidd EAM, Ricketts DNJ *et al* (1993) Occlusal caries diagnosis: a changing challenge for clinicians and epidemiologists. *Journal of Dentistry* **21**: 323-31
- Kidd EM (1989) Caries diagnosis within restored teeth. *Operative Dentistry* **14**: 149-58
- Kidd EM and O'Hara J (1990) The caries status of occlusal amalgams with marginal defects. *Journal of Dental Research* **68**: 1275-7
- Kilpatrick NM, Murray JJ and McCabe JF (1995) The use of a reinforced glass-ionomer cermet for the restoration of primary molars: a clinical trial. *British Dental Journal* **179**: 175-79
- Kilpatrick NM, Murray JJ and McCabe JF (1996) A clinical comparison of a light cured glass ionomer sealant restoration with a composite sealant restoration. *Journal of Dentistry* **24**: 399-405
- Kiremitci A and Bolay S (2003) A 3-year clinical evaluation of gallium restorative alloy. *Journal of Oral Rehabilitation* **30**: 664-7
- Klausner L and Charbeneau G (2006) Amalgam restorations: A cross-sectional survey of placement and replacement. *Journal of the Michigan Dental Association* **67**: 249-52
- Klausner LH and Charbeneau GT (1985) Amalgam restorations: a cross sectional survey of placement and replacement. *Journal of the Michigan Dental Association* **67**: 249-52
- Klausner LM, Green TG and Charbeneau GT (1987) Placement and replacement of amalgam restorations: a challenge for the profession. *Operative Dentistry* **12**: 105-12

- Knapp J (1868) Hidden dental caries. *American Dental Association Transcript* **8**: 108-12
- Knibbs PJ (1992) The clinical performance of a glass polyalkenoate (glass ionomer) cement used in a "sandwich" technique with a composite resin to restore Class II cavities. *British Dental Journal* **172**: 103-7
- Knibbs PJ and Smart ER (1992) The clinical performance of a posterior composite resin restorative material, Heliomolar R.O.: 3-year report. *Journal of Oral Rehabilitation* **19**: 231-7
- Knibbs PJ, Plant CG, Shovelton DS *et al* (1987) An evaluation of a lathe-cut high-copper amalgam alloy. *Journal of Oral Rehabilitation* **14**: 465-7
- Kolker JL, Damiano PC, Jones MP *et al* (2004) The timing of subsequent treatment for teeth restored with large amalgams and crowns: factors related to the need for subsequent treatment. *Journal of Dental Research* **83**: 854-8
- Krejci I, Lieber CM and Lutz F (1995) Time required to remove totally bonded tooth coloured posterior restorations and related tooth substance loss. *Dental Materials* **11**: 34-40
- Kreulen CM, Tobi H, Gruythuysen RJM *et al* (1998) Replacement risk of amalgam treatment modalities: 15 year results. *Journal of Dentistry* **26**: 627-32
- Landis J R and Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**: 159-74
- Lange JP (1991) Clinical markers of periodontal disease. In Johnson NW (ed): Risk markers for Oral Disease, vol 3. Periodontal Disease, Markers of

Disease Susceptibility and Activity. Cambridge. Cambridge University Press, pp 179

- Lavelle CLB (1976) A cross-sectional longitudinal survey into the durability of amalgam restorations. *Journal of Dentistry* **4**: 139-43
- Lazarchik DA, Firestone AR, Heaven TJ *et al* (1995) Radiographic evaluation of occlusal caries: effect of training and experience. *Caries Research* **29**: 355-8
- Leifler E and Varpio M (1981) Proximocclusal composite restorations in primary molars: a two-year follow- up. *Journal of Dentistry for Children* **48**: 411-6
- Leinfelder KF (1985) Evaluation of clinical wear of posterior composite resins. In: Posterior composite resin dental restorative materials, edited by Vanherle G, The Netherlands: Peter Szulc Publ. Co.
- Leinfelder KF, Barkmeier WW and Goldberg AJ (1983) Quantitative wear measurements of posterior composite resins. *Journal of Dental Research* **62(Spec Issue)**: 671
- Leinfelder KF, Sluder TB, Stockwell CL *et al* (1975) Clinical evaluation of composite resins as anterior and posterior restorative materials. *Journal of Prosthetic Dentistry* **33**: 407-16
- Leinfelder KF, Sluder TB, Santos JFF *et al* (1980) Five-year Clinical Evaluation of Anterior and Posterior Restorations of Composite Resin. *Operative Dentistry*. **55**: 57-65
- Leinfelder KF, Wilder AD and Teixeira LC (1986) Wear rates of posterior composite resins. *Journal of the American Dental Association* **112**: 829-33

- Letzel H (1989) Survival rates and reasons for failure of posterior composite restorations in multicentre clinical trial. *Journal of Dentistry* **17**: 10-7
- Letzel H, Aardening CJ, Fick JM *et al* (1978) Corrosion, marginal fracture, and creep of amalgam restorations: a two-year clinical study. *Operative Dentistry* **3**: 82-92
- Letzel H, van 'T Hof MA *et al.* (1989) Failure, survival and reasons for replacement of amalgam restorations. Chicago, Quintessence.
- Letzel H, van 'T Hof M and Vrijhoef M (1987) The influence of the condensation instrument on the clinical behaviour of amalgam restorations. *Journal of Oral Rehabilitation* **14**: 133-8
- Levin KA (2006) Study design IV. Cohort studies. *Evidence Based Dentistry* **7**: 51-2
- Lidums A, Wilkie R and Smales RJ (1993) Occlusal glass ionomer cermet, resin sandwich and amalgam restorations: a 2-year clinical study. *American Journal of Dentistry* **6**: 185-8
- Lotzkar L, Johnson DW *et al* (1971) Experimental program in expanded functions for dental assistants: phase 1 base line and phase 2 training. *Journal of the American Dental Association* **82**: 101
- Lucarotti PSK, Holder RL and Burke FJT (2005a) Analysis of an administrative database of half a million restorations over 11 years. *Journal of Dentistry* **33**: 791-803
- Lucarotti PSK, Holder RL and Burke FJT (2005b) Outcome of direct restorations placed within the general dental services in England and Wales (Part3): Variation by dentist factors. *Journal of Dentistry* **33**: 827-35

- Lundin SA and Koch G (1989) Class I and II composite resin restorations: 4-year clinical follow up. *Swedish Dental Journal* **13**: 217-27
- Lundin SA, Andersson B, Koch G and Rasmusson CG (1990) Class II composite resin restorations: a three-year clinical study of six different posterior composites. *Swedish Dental Journal* **1990 14**: 105-14
- Lussi A, Firestone A, Schoenberg V *et al* (1995) In vivo diagnosis of fissure caries using a new electrical resistance monitor. *Caries Research* **29**: 81-7
- Lynch CD, McConnell RJ and Wilson NHF (2007) trends in the placement of posterior composite in dental schools. *Journal of International Dental Education* **71**: 430-4
- Lucarotti PSK, Holder RL and Burke FJT (2005a) Analysis of an administrative database of half a million restorations over 11 years. *Journal of Dentistry* **33**: 791-803
- Lucarotti PSK, Holder RL and Burke FJT (2005b) Outcome of direct restorations placed within the general dental services in England and Wales (Part1): Variation by type of restoration and re-intervention. *Journal of Dentistry* **33**: 805-15
- Lundin SA and Koch G (1989) Class I and II composite resin restorations: 4-year clinical follow up. *Swedish Dental Journal* **13**: 217-27
- Lussi A (1993) Comparison of different methods for the diagnosis of fissure caries without cavitation. *Caries Research* **27**: 409-16
- MacInnis WA, Ismail A and Brogan H (1991) Placement and replacement of restorations in a military population. *Journal of the Canadian Dental Association* **57**: 227-31

- Mahler DB and Adey J (1977) Microprobe analysis of high copper amalgam. *Journal of Dental Research* **3**: 379-84
- Mahler DB and Marantz R (1979) The effect of the operator on the clinical performance of amalgam. *Journal of the American Dental Association* **99**: 38-41
- Mahler DB, Engle JH and Adey JD (1990) Effect of Pd on the clinical performance of amalgam. *Journal of Dental Research* **69**: 1759-61
- Mahler DB, Terkla LG *et al* (1973) Marginal fracture of amalgam restorations. *Journal of Dental Research* **52**: 823-7
- Mahler DB, Terkla LG, Van Eysden J *et al* (1970) Marginal Fracture vs Mechanical Properties of Amalgam. *Journal of Dental Research* **49**: 1452-57
- Mahmood S and Smales RJ (1994) Longevity of dental restorations in selected patients from different practice environments. *Australian Dental Journal* **39**: 15-7
- Mahmood S, Chohan AN, Al-Jannakh M *et al* (2004) Placement and replacement of dental restorations. *Journal of the College of Physicians and Surgeons – Pakistan* **14**: 589-92
- Mair LH (1995) Wear patterns in two amalgams and three posterior composites after 5 years' clinical service. *Journal of Dentistry* **23**: 107-12
- Mair LH (1998) Ten-year clinical assessment of three posterior resin composites and two amalgams. *Quintessence International* **29**: 483-90
- Manhart J, Chen HY, Hamm G *et al* (2004) Review of the clinical survival of direct and indirect restorations in posterior teeth of the permanent dentition. *Operative Dentistry* **29**: 481-508

- Manhart J and Hickel r (2001) Longevity of restorations. In: Wilson NHF, Roulet J-F and Fuzzi M (eds) *Advances in operative dentistry- Challenges of the future*. Vol 2. Quintessence, Chicago, pp237-304
- Martin JA and Bader JD (1997) Five-year outcomes for teeth with large amalgams and crowns. *Operative Dentistry* **22**: 72-8
- Marynuik GA (1984) In search of treatment longevity- a 30-year perspective. *Journal of the American Dental Association* **109**: 739-44
- Marynuik GA (1990a) Practice variation: learned and socioeconomic factors. *Advances in Dental Research* **4**: 19-24
- Marynuik GA (1990b) Replacement of amalgam restorations that have marginal defects: variation and cost implications. *Quintessence International* **21**: 111-3
- Marynuik GA and Brunson WD (1989) When to replace faulty-margin amalgam restorations, a pilot study. *General Dentistry* **37**: 463-7
- Marynuik GA and Kaplan S (1986) Longevity of restorations: survey results of dentists' estimates and attitudes. *Journal of the American Dental Association* **112**: 39-45
- Matis BA, Carlson T, Cochran M *et al* (1991). How finishing affects glass ionomers. Results of a five-year evaluation. *Journal of the American Dental Association*. **122**: 43-6
- Matis BA, Cochran M and Carlson T (1996) Longevity of glass-ionomer restorative materials: results of a 10-year evaluation. *Quintessence International* **27**: 373-82

- Matsson L, Ryge G, Weidemanis C *et al* (1982) Adaption of Dispersion and Traditional amalgams with reference to plasticity: A Clinical Comparison. *Journal of Dental Restoration* **61**: 1172-5
- Maupomé G (1998) A comparison of senior dental students and normative standards with regard to caries assessment and treatment decisions to restore occlusal surfaces of permanent teeth. *Journal of Prosthetic Dentistry* **79**: 596-603
- Maupomé G and Sheiham A (1998) Criteria for restoration replacement and restoration life-span estimates in an educational environment. *Journal of Oral Rehabilitation* **25**: 896-901
- Mausner JS and Kramer S (1985) Epidemiology: An introductory text 2nd edition. WB Saunders Co, the Curtis Center, Independence Square West, Philadelphia, P! 19106 ISBN 0-7216-6181-5
- McAndrew R and Longbottom C (1993) Magnification as an aid to caries diagnosis: an *in-vitro* study. *Journal of Dental Research* **72**: 708
- McAndrew R and Longbottom C (1994) The effect of soft tissue equivalent material in caries diagnosis- *in vitro* *Journal of Dental Research* **73**: 337
- McComb D and Tam LE (2001) Diagnosis of occlusal caries: Part 1. Conventional methods. *Journal of the Canadian Dental Association* **67**: 454-7
- Merrett MC and EldertonRJ (1984) An in vitro study of restorative dental treatment decisions and dental caries. *British Dental Journal* **157**: 128-33
- Mertz-Fairhurst EJ *et al* (1987) Clinical performance of sealed composite restorations placed over caries compared with sealed and unsealed

amalgam restorations. *Journal of the American Dental Association* **115**: 689-94

- Mileman P and van der Weele LT (1990) Accuracy in radiographic diagnosis: Dutch practitioners and dental caries. *Journal of Dentistry* **18**: 130-6
- Mileman P, Mulder H and van der Weele L (1992) Factors influencing the likelihood of successful decisions to treat dentin caries from bitewing radiographs. *Community Dentistry and Oral Epidemiology* **20**: 175-80
- Milgrom P, Kiyak HA, Conrad D *et al* (1981) A study of treatment planning: periodontal services for the elderly. *Journal of Dental Education* **45**: 522-8
- Millar BJ, Robinson PB and Inglis AT (1997) Clinical evaluation of an anterior hybrid composite resin over 8 years. *British Dental Journal* **182**: 26-30
- Mills DC and Hollis S (1997) training for dental screening using clinical photographs. *Journal of Community Dental Health* **14**: 245-7
- Mjör IA (1981) Placement and replacement of restorations. *Operative Dentistry* **6**: 49-54
- Mjör IA (1985) The frequency of secondary caries at various anatomical locations. *Operative Dentistry* **10**: 88-92
- Mjör IA (1986) Clinical assessments of amalgam restorations. *Operative Dentistry* **77**: 55-62
- Mjör IA (1989) Amalgam and composite resin restorations: longevity and reasons for replacement. Quality evaluation of dental restorations. Criteria for placement and replacement. K. J. Anusavce. Chicago, Quintessence: pp 61-72

- Mjör IA (1993) Direct posterior filling materials. In: Vanherle G, Derange M, Willems G, editors. State-of-the-art on direct posterior filling materials and dentine bonding. Leuven: van der Poorten, 1993: pp1-13
- Mjör IA (1997a) Selection of restorative materials in general dental practice in Sweden. *Acta Odontologica Scandinavica* **55**: 53-7
- Mjör IA (1997b) The reasons for replacement and the age of failed restorations in general dental practice. *Acta Odontologica Scandinavica* **55**: 58-63
- Mjör IA (1999) Selection of restorative materials in general dental practice in Sweden. *Acta Odontologica Scandinavica* **57**: 257-62
- Mjör IA (2004) A recurring problem: research in restorative dentistry. But there is a light at the end of the tunnel. *Journal of Dental Research* **83**: 92
- Mjör IA and Gordon VV (2002) Failure, repair, refurbishing and the longevity of restorations. *Operative Dentistry* **27**: 528-34
- Mjör IA and Haugen E (1976) Clinical evaluation of amalgam restorations. *Scandinavian Journal of Dental Research* **84**: 333-7
- Mjör IA and Jokstad A (1993) A Five-year study of Class II restorations in permanent teeth using amalgam, glass polyalkenoate (ionomer) cement and resin-based composite materials. *Journal of Dentistry* **21**: 338-43
- Mjör IA and Medina JE (1993) Reasons for placement, replacement and age of gold restorations in selected practices. *Operative Dentistry* **18**: 82-7
- Mjör IA and Moorehead J (1993) Selection of restorative dental materials. Reasons for failures and longevity of restorations in Florida. *Journal of Dental Research* **18**: 82-7

- Mjör IA and Moorhead E (2000) Reasons for replacement of restorations in permanent teeth in general dental practice. *International Dental Journal* **50**: 361-8
- Mjör IA and Qvist V (1997) Marginal failures of amalgam and composite restorations. *Journal of Dentistry* **7**: 25-30
- Mjör IA and Ryge G (1981) Comparison of techniques for the evaluation of marginal adaptation of amalgam restorations. *International Dental Journal* **31**: 1-5
- Mjör IA and Toffenetti F (1992) Placement and replacement of amalgam restorations in Italy. *Operative Dentistry* **17**: 70-3
- Mjör IA and Toffenetti F (1992) Placement and replacement of resin based composite restorations in Italy. *Operative Dentistry* **17**: 82-5
- Mjör IA and Toffenetti F (2000) Secondary caries: a literature review with case reports. *Quintessence International* **31**: 165-79
- Mjör IA and Um CM (1993) Survey of amalgam and composite restorations in Korea. *International Dental Journal* **43**: 311-6
- Mjör IA and Wilson NHF (1997) The relative cost of different restorations in the UK. *British Dental Journal* **182**: 286-9
- Mjör IA, Dahl JE and Moorhead JE (2000) The age of restorations at placement in permanent teeth in general dental practice. *Acta Odontologica Scandinavica* **58**: 97-101
- Mjör IA, Dahl JE and Moorhead JE (2002) Placement and replacement of restorations in primary teeth. *Acta Odontologica Scandinavica* **60**: 25-28
- Mjör IA, Gordan VV, Abuhanna A et al (2005) Research in general dental practice. *Acta Odontologica Scandinavica* **63**: 1-9

- Mjör IA, Jokstad A and Qvist V (1990) Longevity of posterior restorations. *International Dental Journal* **40**: 11-7
- Mjör IA, Moorhead JE and Dahl JE (2000) Reasons for replacement of restorations in permanent teeth in general dental practice. *International Dental Journal* **50**: 360-6
- Mjör IA, Shen C, Eliason ST *et al* (2002) Placement and replacement of restorations in General Dental Practice in Iceland. *Operative Dentistry* **27**: 117-23
- Moffa JP (1989) The longevity and reasons for replacement of amalgam alloys. *Journal of Dental Research* **68 (Special Issue)**: 188 (Abstr. 56)
- Moher D, Schulz KF and Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet* **357**: 1191-4
- Moore DL and Stewart JL (1967) Prevalence of defective dental restorations. *Journal of Prosthetic Dentistry* **17**: 372-8
- Mormann W and Krejci I (1992) Computer-designed inlays after 5 years in situ: clinical performance and scanning electron microscopic evaluation. *Quintessence International* **23**: 109-15
- Morris ME, Braham RL, Schmutz JR *et al* (1981) A clinical and laboratory study comparing three amalgam alloys of random particle-size, mixed phase with one of conventional regular lathe-cut particles. *Acta de Odontologia Pediatrica* **2**: 41-5
- Munn P and Drever E (1996) Using questionnaires in small-scale research: a teacher's guide. *Scottish Council for research in Education*. 15 St John Street, Edinburgh EH8 8JR

- Natkin E and Guild RE (1967) An evaluation of preclinical laboratory performance: a systematic study. *Journal of Dental Education* **31**: 152-61
- Navarro MF, Franco EB, Bastos PA *et al* (1996) Clinical evaluation of gallium alloy as a posterior restorative material. *Quintessence International* **3**: 115-8
- Nelson GV, Osborne JW, Gale EN *et al* (1980) A three-year clinical evaluation of composite resin and a high copper amalgam in posterior primary teeth. *Journal of Dentistry for Children* **47**: 414-8
- Neo J and Chew CL (1996) Direct tooth-colored materials for noncarious lesions: a 3-year clinical report. *Quintessence International* **27**: 183-8
- Noak MJ, van Meerbeek B, Roulet JF *et al* (1995) Marginal behaviour of 10 different tooth-coloured inlay systems in vivo. *Journal of Dental Research* (Spec. Iss.) **74**: 165 (Abstract 1227)
- Noar SJ and Smith BG (1990) Diagnosis of caries and treatment decisions in approximal surfaces of posterior teeth in vitro. *Journal of Oral Rehabilitation* **17**: 209-18
- Nordbo H and Lyngstadaas SP (1992) The clinical performance of two groups of functioning class-II cast gold alloys. *Acta Odontologica Scandinavica* **22**: 857-64
- Nordbo H, Leirskar J *et al* (1998) Saucer-shaped cavity preparations for posterior approximal resin composite restorations: Observations up to 10 years. *Quintessence International* **29**: 5-11
- Norman RD and Wilson NH (1988) Three-year findings of a multiclinical trial for a posterior composite. *Journal of Prosthetic Dentistry* **59**: 577-83

- Norman RD, Wright JS, Rydberg RJ *et al* (1990) A 5-year study comparing a posterior composite resin and an amalgam. *Journal of Prosthetic Dentistry* **64**: 523-9
- Nuttall N M and Elderton RJ (1983) The nature of restorative dental treatment decisions. *British Dental Journal* **154**: 363-5
- Nyvad B, Machiulskiene V and Baelum V (1999) Reliability of a new caries diagnostic system differentiating between active and inactive caries lesions. *Caries Research* **23**: 252-60
- Oldenburg TR, Vann WF and Dilley DC (1987) Comparison of composite and amalgam in posterior teeth of children. *Dental Materials* **3**: 182-6
- Oldenburg TR, Vann WF, Jr. and Dilley DC (1985) Composite restorations for primary molars: two-year results. *Pediatric Dentistry* **7**: 96-103
- Olmez A and Ulusu T (1995) Bond strength and clinical evaluation of a new dentinal bonding agent to amalgam and resin composite. *Quintessence International* **26**: 785-93
- Opdam NJ, Feilzer AJ, Roeters JJ *et al* (1998) Class I occlusal composite resin restorations: in vivo post-operative sensitivity, wall adaptation and microleakage. *American Journal of Dentistry* **11**: 319-27
- Osborne JW (2006) Amalgam: Dead or alive. *Dental Journal* **33**: 94-8
- Osborne JW and Friedman SJ (1986) Clinical evaluation of marginal fracture of amalgam restorations: one-year report. *Journal of Prosthetic Dentistry* **55**: 335-9

- Osborne JW and Gale EN (1974a) A two-, three-, and four-year follow-up of a clinical study of the effect of trituration on amalgam restorations. *Journal of the American Dental Association* **88**: 795-7
- Osborne JW and Gale EN (1974b) Long-term follow-up of clinical evaluations of lathe cut versus spherical amalgam. *Journal of Dental Research* **53**: 1204-7
- Osborne JW, Binon PP *et al* (1980) Dental amalgam: Clinical behaviour up to eight years. *Operative Dentistry* **5**: 24-8
- Osborne JW, Gale EN and Fergusson GW (1973) One-year and two-year clinical evaluations of a composite resin vs. amalgam. *Journal of Prosthetic Dentistry* **30**: 795-800
- Osborne JW, Norman RD and Gale EN (1991) A 14-year clinical assessment of 12 amalgam alloys. *Quintessence International* **22**: 857-64
- Osborne JW, Norman RD Swartz ML *et al* (1978) In vivo comparison of a composite resin and its radiopaque counterpart. *Journal of Prosthetic Dentistry* **39**: 406-8
- Osborne, JW, Phillips WR *et al* (1976) Three-year clinical comparison of three amalgam alloy types emphasizing an appraisal of the evaluation methods used. *Journal of the American Dental Association* **93**: 784-9
- Pallensen U and Qvist V (2003) Composite resin fillings and inlays. An 11-year evaluation. *Clinical Oral Investigations* **7**: 71-9
- Palotie U and Vehkalahti M (2002) Reasons for replacement and the age of failed restorations in posterior teeth of young Finnish adults. *Acta Odontologica Scandinavica* **60**: 325-329

- Paquette DE, Vann WF, Oldenburg TR *et al* (1983) Modified cavity preparations for composite resins in primary molars. *Pediatric Dentistry* **5**: 246-251
- Paterson N (1984) The longevity of restorations: a study of 200 regular attenders in a general dental practice. *British Dental Journal* **157**: 857-64
- Phillips RW, Avery DR, Mehra R *et al* (1972) Observations on a composite resin for Class II restorations. Two-year report. *Journal of Prosthetic Dentistry* **28**: 164-9
- Phillips RW, Avery DR, Mehra R *et al.* (1973) Observations on a composite resin for Class II restorations; three-year report. *Journal of Prosthetic Dentistry* **30**: 891-7
- Pimenta LAF, de Lima Navarro MF *et al* (1995) Secondary caries around amalgam restorations. *Journal of Prosthetic Dentistry* **74**: 219-22
- Pink FE, Minden NJ *et al* (1994) Decisions of practitioners regarding replacement of amalgam and composite restorations in general dental practice settings. *Operative Dentistry* **19**: 127-32
- Pitts NB (1997) Diagnostic tools and measurements-impact on appropriate care *Community Dentistry and Oral Epidemiology* **25**: 24-35
- Pitts NB and Fyffe HE (1987) The effect of varying diagnostic thresholds upon clinical caries data for a low prevalence group. *Journal of Dental Research* **67**: 592-6
- Plasmans PJJM, Creugers NHJ and Mulder J (1998) Long-term survival of extensive amalgam restorations. *Journal of Dental Research* **77**: 453-60
- Poorterman JH, Aartman IH and Kalsbeek H (1999) Underestimation of the prevalence of approximal caries and inadequate restorations in a clinical

epidemiological study. *Community Dentistry and Oral Epidemiology* **27**: 331-7

- Poulsen E A (1987) Method for training and checking interrater agreement for a patient classification study. *Nursing Management* **18**: 72-80
- Poulsen E, Bille J and Rugg-Gunn AJ (1980) Evaluation of a calibration trial to increase interexaminer reliability of radiographic diagnosis of approximal carious lesions. *Community Dentistry and Oral Epidemiology* **8**: 135-8
- Powell LV, Johnson GH and Gordon GE (1995) Factors associated with clinical success of cervical abrasion/erosion restorations. *Operative Dentistry* **20**: 7-13
- Pretty IA and Maupomé G (2004a) A closer look at diagnosis in clinical dental practice: Part 1. Reliability, validity, specificity and sensitivity of diagnostic procedures. *Journal of the Canadian Dental Association* **70**: 251-5
- Pretty IA and Maupomé G (2004b) A closer look at diagnosis in clinical dental practice: Part 2. Using predictive values and receiver operating characteristics in assessing diagnostic accuracy. *Journal of the Canadian Dental Association* **70**: 313-6
- Pretty IA and Maupomé G (2004c) A closer look at diagnosis in clinical dental practice: Part 3. Effectiveness of radiographic diagnostic procedures. *Journal of the Canadian Dental Association* **70**: 388-94
- Pretty IA and Maupomé G (2004d) A closer look at diagnosis in clinical dental practice: Part 4. Effectiveness of nonradiographic diagnostic procedures and devices in dental practice. *Journal of the Canadian Dental Association* **70**: 470-4

- Pretty IA and Maupomé G (2004e) A closer look at diagnosis in clinical dental practice: Part 5. Emerging technologies for caries detection and diagnosis. *Journal of the Canadian Dental Association* **70**: 540a-g
- Pretty IA and Maupomé G (2004f) A closer look at diagnosis in clinical dental practice: Part 6. Emerging technologies for detection and diagnosis of noncaries dental problems. *Journal of the Canadian Dental Association* **70**: 621-6
- Pretty IA, Ingram G, Agalamanyi EA *et al* (2003) The use of flourescein-enhanced quantitative light-induced fluorescence to monitor de- and re-mineralisation of in vitro root caries. *Journal of Oral Rehabilitation* **30**: 1151-6
- Qualtrough AJ and Wilson NH (1996) A 3-year clinical evaluation of a porcelain inlay system. *Journal of Dentistry* **24**: 317-23
- Qvist V, Qvist J *et al* (1990) Placement and longevity of tooth-colored restorations in Denmark. *Acta Odontologica Scandinavica* **48**: 305-11
- Qvist V, Strom C and Thylstrup A (1985) Two-year assessment of anterior resin restorations inserted with two acid-etch restorative procedures. *Scandinavian Journal of Dental Research* **93**: 343-50
- Qvist V, Thylstrup A and Mjör IA (1986a) Restorative treatment pattern and longevity of amalgam restorations in Denmark. *Acta Odontologica Scandinavica* **44**: 343-9
- Qvist V, Thylstrup A and Mjör IA (1986b) Restorative treatment pattern and longevity of resin restorations in Denmark. *Acta Odontologica Scandinavica* **44**: 351-6

- Qvist, J, Qvist V *et al* (1990a) Placement and longevity of amalgam restorations in Denmark. *Acta Odontologica Scandinavica* **48**: 297-303
- Raadal M (1978) Follow-up study of sealing and filling with composite resins in the prevention of occlusal caries. *Community Dentistry & Oral Epidemiology* **6**: 176-80
- Rasmusson CG, Kholer B and Odman P (1998) A 3-year clinical evaluation of two composite resins in classII cavities. *Acta Odontologica Scandinavica* **56**
- Rasmusson CG and Lundin SA (1995) Class II restorations in six different posterior composite resins: five-year results. *Swedish Dental Journal* **19**: 173-82
- Redisch M (1978) Physician involvement in hospital decision making. Hospital cost containment: selected notes for future policy in Zubkoff M, Raskin E and Hanft R. New york, Milbank Memorial Fund: 217-43.
- Richardson AS and Boyd MA (1973) Replacement of silver amalgam restorations by 50 dentists during 246 working days. *Journal of the Canadian Dental Association* **39**: 556-9
- Richardson R, Treasure E and Sheldon T (1999) On the evidence: Dental restoration. Health Service Journal. Apr 8; 109(5649):28-9. NHS Centre for Reviews and Dissemination
- Ricketts D, Kidd EM *et al* (1995) Clinical and radiographic diagnosis of occlusal caries: a study in vitro. *Journal of Oral Rehabilitation* **22**: 15-20
- Robbins JW and Summitt JB (1988) Longevity of complex amalgam restorations. *Operative Dentistry* **13**: 54-7

- Robertello FJ and Pink FE (1997) The effect of a training program on the reliability of examiners evaluating amalgam restorations. *Operative Dentistry* **22**: 57-65
- Roberts MW, Broring CL and Moffa JP (1985) Two-year clinical evaluation of a proprietary composite resin for the restoration of primary posterior teeth. *Pediatric Dentistry* **7**: 14-8
- Roberts MW, Folio J, Moffa JP *et al* (1992) Clinical evaluation of a composite resin system with a dentin bonding agent for restoration of permanent posterior teeth: a 3-year study. *Journal of Prosthetic Dentistry* **67**: 301-6
- Robinson AA, Rowe AH and Maberley ML (1988) A three-year study of the clinical performance of a posterior composite and a lathe cut amalgam alloy. *British Dental Journal* **164**: 248-52
- Robinson AD (1971) The life of a filling. *British Dental Journal* **130**: 206-8
- Roulet JF (1988) The problems with substituting composite resins for amalgam: a status report on posterior composites. *Journal of Dental Research* **68**: 101-13
- Roulet JF (1997) Benefits and disadvantages of tooth-coloured alternatives to amalgam. *Journal of Dentistry* **25**: 459-73
- Roulet JF, Reich T, Blunck U *et al* (1989) Quantitative margin analysis in the scanning electron microscope. *Scanning Electron Microscope* **3**: 147-59
- Rudolphy MP, van Amerongen JP, Penning C *et al* (1995) Grey discolouration and marginal fracture for the diagnosis of secondary caries in

molars with Occlusal amalgam restorations: an in vitro study. *Caries Research* **29**: 371-6

- Rugg-Gunn AJ and Holloway PJ (1974) Methods of measuring the reliability of caries prevalence and incremental data. *Community Dentistry & Oral Epidemiology* **2**: 287-94
- Rydinge E, Goldberg J, Sanchez L *et al* (1981) Clinical evaluation of high copper amalgam restorations. *Journal of Oral Rehabilitation* **8**: 465-72
- Ryge G and Snyder M (1973) Evaluating the quality of dental restorations. *Journal of the American Dental Association* **87**: 369-77
- Ryge G, Cvar JF, Micik RE *et al* (1974) Clinical studies of restorative materials. 1: zinc vs. non-zinc alloy. *Journal of Dental Research* **53**: 61
- Rykke M (1992) Dental materials for posterior restorations. *Endodontic Dentistry and Traumatology* **8**: 139-48
- Rytömaa I, Jarvinen V and Jarvinen I (1979) Variation in caries recording and restorative treatment plan among university teachers. *Community Dentistry and Oral Epidemiology* **7**: 335-9
- Rytömaa I, Murtomaa H *et al* (1983) Clinical assessment of amalgam fillings. *Community Dentistry and Oral Epidemiology* **12**: 169-72
- Saleh N *et al* (1992) One-year clinical evaluation of an anterior composite resin. *Quintessence International* **23**: 559-67
- Sano H, Kanemura N, Burrow MF *et al* (1998) Effect of operator variability on dentin adhesion: students versus dentists. *Dental Materials* **17**: 51-8
- Scheer B (1975) The restoration of injured anterior teeth in children by etch-retained resin. A longitudinal study. *British Dental Journal* **139**: 465-8

- Schonfeld HK (1967) The quality of dental care in community programs. *Journal of Public Health Dentistry* **27**: 270
- Schonfeld HK (1969) Peer review of quality of dental care. *Journal of the American Dental Association* **79**: 1376
- Schuller AA and Holst D (1998) Changes in the oral health of adults from Trondelag, Norway, 1973-1983-1994. *Community Dentistry and Oral Epidemiology* **26**: 201-208
- Scruggs RR, Daniel SJ, Larkin AA *et al* (1989) Effects of specific criteria and calibration on examiner reliability. *Journal of Dental Hygiene* **63**: 125-9
- Setcos JC, Staninec M and Wilson NH (1999) A two-year randomized, controlled clinical evaluation of bonded amalgam restorations. *Journal of Adhesive Dentistry* **1**: 323-31
- Shaw L and Murray JJ (1975) Inter-examiner and intra-examiner reproducibility in clinical and radiographic diagnosis. *International Dental Journal* **25**: 280-8
- Sjögren G, Molin M, van Dijken J *et al* (1995) Ceramic inlays (Cerec) cemented with either a dual-cured or a chemically cured composite resin luting agent. A 2-year clinical study. *Acta Odontologica Scandinavica*. **53**: 325-30
- Skogedal O and Helöe LA (1979) Clinical quality of amalgam restorations. *Scandinavian Journal of Dental Research* **87**: 459-61
- Slack GL, Jackson D, James PMC *et al* (1958) A clinical investigation into the variability of dental caries diagnosis. *British Dental Journal* **104**: 399-404

- Smales RJ (1982) Fissure sealants versus amalgams: clinical results over five years. *Journal of Dentistry* **10**: 95-102
- Smales RJ (1983) Incisal angle adhesive resins: a 5-year clinical survey of two materials. *Journal of Oral Rehabilitation* **10**: 19-24
- Smales RJ (1991) Longevity of cusp-covered amalgams: survival after 15 years. *Operative Dentistry* **16**: 17-20
- Smales RJ and Creaven PJ (1979) Evaluation of clinical methods for assessing the surface roughness of restorations. *Journal of Prosthetic Dentistry* **42**: 45-52
- Smales RJ and Creaven PJ (1985) Evaluation of three clinical methods for assessing amalgam and resin restorations. *Journal of Prosthetic Dentistry* **54**: 340-6
- Smales RJ and Hawthorne WS (1996) Long-term survival and cost-effectiveness of five dental restorative materials used in various classes of cavity preparations. *International Dental Journal* **46**: 126-30
- Smales RJ and Hawthorne WS (1997) Long-term survival of extensive amalgams and posterior crowns. *Journal of Dentistry* **25**: 225-7
- Smales RJ, Gerke DC and White IL (1990) Clinical evaluation of occlusal glass-ionomer, resin and amalgam restorations. *Journal of Dentistry* **18**: 243-9
- Smales RJ, Webster DA *et al* (1991a) Survival predictions of four types of dental restorative materials. *Journal of Dentistry* **19**: 857-64
- Smales, RJ and Hawthorne WS (1996) Long-term survival and cost-effectiveness of five dental restorative materials used in various classes of cavity preparations. *International Dental Journal* **46**: 126-30.

- Smales, RJ, Webster DA and Leppard PI (1991b) Prediction of amalgam restoration longevity. *Journal of Dentistry* **19**: 18-23
- Smith GA and Wilson NH (1979) A visible light-cured composite restorative. A clinical trial. *British Dental Journal* **147**: 185-7
- Söderholm K-JM, Antonson DE and Fischlschweiger W (1989) Correlation between marginal discrepancies at the amalgam / tooth interface and recurrent caries. In: Anusavice KJ (Ed) Quality evaluation of dental restorations. Quintessence Publishing, Chicago, pp 95-108
- Soricelli DA (1970). Controlling quality in the delivery of dental care. 21st National Dental Health Conference, Chicago.
- Stangel I and Barolet RY (1990) Clinical evaluation of two posterior composite resins: two-year results. *Journal of Oral Rehabilitation* **17**: 257-68
- Straffon LH, Corpron RE, Dennison JB *et al* (1983) A clinical evaluation of polished and unpolished amalgams: 18 month results. *Pediatric Dentistry* **5**: 177-82
- Straffon LH, Corpron RE, Dennison JB *et al* (1984) A clinical evaluation of polished and unpolished amalgam: 36-month results. *Pediatric Dentistry* **6**: 220-5
- Striffler DN (1963) Quality control in dentistry is integrity and courage. 14th National Dental Health Conference, Chicago
- Sturdevant JR, Lundeen TF, Slunder TB *et al* (1986) Three-year study of two light cured composite resins. *Dental Materials* **2**: 263-8
- Sturdevant JR, Lunden TF, Sluder Jr. TB *et al* (1988) Five year study of two light-cured posterior composite resins. *Dent Materials* **4**: 105-10

- Summitt JB, Burgess JO, Berry TJ *et al* (2001) The performance of bonded vs pin-retained complex amalgam restorations. A five year clinical evaluation. *Journal of the American Dental Association* **132**: 923-31
- Taylor DF, Turnbull CD and Leinfelder KF (1984) Comparative evaluation of casts for the measurement of composite wear. *Journal of Dental Research* **64**(Spec Issue):293
- Thordrup M, Isidor F and Horsted-Bindslev P (1998) A five-year clinical study of tooth coloured inlays. *Journal of Dental Research* (IADR Abstracts); **77**: 912 (number 2255)
- Tidehag P and Gunne J (1995). A 2-year clinical follow-up study of IPS Empress ceramic inlays. *International Journal of Prosthodontics*. **8**: 456-60
- Todd JE (1995) Children's *dental health in England and Wales*. Office of Population Census and Surveys. Her Majesty's Stationery Office, London pp 329-33
- Tonn EM and Ryge G (1985) Two-year clinical evaluation of light-cured composite resin restorations in primary molars. *Journal of the American Dental Association* **111**: 44-8
- Tonn EM, Ryge G and Chambers DW (1980). A two-year clinical study of a carvable composite resin used as class II restorations in primary molars. *ASDC Journal of Dentistry for Children* **47**: 405-13
- Tveit AB and Espelid I (1986) Radiographic diagnosis of caries and marginal defects in connection with radiopaque composite fillings. *Dental Materials* **2**: 159-62

- Tveit AB and Espelid I (1992) Class II amalgams: interobserver variations in replacement decisions and diagnosis of caries and crevices. *International Dental Journal* **42**: 12-8
- Tveit AB, Espelid I, Erickson RL *et al* (1991) Vertical angulation of the X-ray beam and radiographic diagnosis of secondary caries. *Community Dentistry and Oral Epidemiology* **19**: 333-5
- Tyas MJ (1991) Cariostatic effect of glass ionomer cement: a five-year clinical study. *Australian Dental Journal* **36**:236-9
- Tyas MJ (2005) Placement and replacement of restorations by selected practitioners. *Australian Dental Journal* **50**: 81-9
- Tyas MJ (1984) Three year clinical evaluation of Tenure bonding agent. *Australian Dental Journal* **39**: 188-9
- van Amerongen WE and Eggink CO (1986) The cervical margin of amalgam restorations: a radiographic and clinical assessment. *ASDC Journal of Dentistry for Children* **53**: 177-83
- van Dijken JW (1986) A clinical evaluation of anterior conventional, microfiller, and hybrid composite resin fillings. A 6-year follow-up study. *Acta Odontologica Scandinavica* **44**: 357-67
- van Dijken JW (1991) A six year follow-up of three dental alloy restorations with different copper contents. *Swedish Dental Journal* **15**: 259-64
- van Dorp CS, Exterkate RA *et al* (1988) The effect of dental probing on subsequent enamel demineralization. *ASDC Journal of Dentistry for Children* **55**: 343-7

- Van Meerbeek B, Peumans M and Gladys S (1996) Three-year clinical effectiveness of four total-etch dentinal adhesive systems in cervical lesions. *Quintessence International* **27**: 775-84
- Van Nieuwenhuysen JP, D'Hoore W, Carvalho J *et al* (2003) Long-term evaluation of extensive restorations in permanent teeth. *Journal of Dentistry* **31**: 395-405
- Verdonschotsh EH, Bronkhorst EM, Burgersdijk RCS *et al* (1992) Performance of some diagnostic systems in examinations for small occlusal caries. *Caries Research* **26**: 59-64
- Von Korff M, Howard J, Truelove E *et al* (1988) Temporomandibular disorders: variations in clinical practice. *Medical Care* **26**: 307-14
- Wahl MJ, Schmitt MM, Overton DA *et al* (2004) Prevalence of cusp fractures in teeth restored with amalgam and resin-based composite. *Journal of the American Dental Association* **135**: 1127-32
- Walls AW, Murray JJ and McCabe JF (1988) The management of occlusal caries in permanent molars. A clinical trial comparing a minimal composite restoration with an occlusal amalgam restoration. *British Dental Journal* **164**: 288-92
- Walls AWG, Wallwork MA *et al* (1985) The longevity of occlusal amalgam restorations in first permanent molars of child patients *British Dental Journal* **158**: 133-6
- Walsh TF and Saxby MS (1989) Inter- and intra-examiner variability using standard and constant force periodontal probes. *Journal of Dental Research* **65**: 432-6

- Warren JJ, Steven ML and Kanellis MJ (2002) Dental caries in the primary dentition: assessing prevalence of cavitated and noncavitated lesions. *Journal of Public Health Dentistry* **62**: 109-14
- Weaver JR and Saeger W (1984) A rater training program in removable prosthodontics. *Journal of Dental Education* **48**: 251-7
- Weintraub JA, Stearns SC, Burt B *et al* (1993) A Retrospective Analysis of the Cost-Effectiveness of Dental Sealants in a Children's Health Centre. *Social Science and Medicine* **36**: 1483-93
- Welbury RR and Murray JJ (1990) A clinical trial of the glass-ionomer cement-composite resin "sandwich" technique in Class II cavities in permanent premolar and molar teeth. *Quintessence International* **21**: 507-12
- Welbury RR, Walls AW, Murray JJ *et al* (1991) The 5-year results of a clinical trial comparing a glass polyalkenoate (ionomer) cement restoration with an amalgam restoration. *British Dental Journal* **170**: 177-81
- Welbury RR, Walls AWG, Murray JJ *et al* (1990) The management of occlusal caries in permanent molars. A 5-year clinical trial comparing a minimal composite with an amalgam restoration. *British Dental Journal* **169**: 361-6
- Wendt SL and Leinfelder KF (1992) Clinical evaluation of Clearfil photoposterior: 3-year results. *American Journal of Dentistry* **5**: 121-5
- Wendt SL and Leinfelder KF (1994). Clinical evaluation of a posterior resin composite: 3-year results. *American Journal of Dentistry* **7**: 207-11

- Wendt SL and Leinfelder KF (1990) The clinical evaluation of heat-treated composite resin inlays. *Journal of the American Dental Association* **120**: 177-181
- Wenneberg JE (1984) Dealing with medical practice variations; a proposal for action. *Health Affairs* **3**: 6-32
- Wenzel A, Fejerskov O, Kidd E *et al* (2000) Depth of occlusal caries assessed clinically, by conventional film radiographs, and by digitized, processed radiographs. *Caries Research* **24**: 327-33
- Whitehead SA and Wilson NHF (1992) Restorative decision-making behaviour with magnification. *Quintessence International* **23**: 667-71
- Widmer CG (1995) Physical characteristics associated with temperomandibular disorders. In Sessle BJ, Bryant PS, Dionne RA (eds): *Temperomandibular disorders and related pain conditions, progress in pain resolution and management, volume 4*. Seattle, IASP Press, pp 161-74
- Widström E and Forss H (1991) Safety of dental restorative materials: a survey of dentists' attitudes. *Proceedings of the Finnish Dental Society* **87**: 351-7
- Wilder AD, Bayne SC, May KN *et al* (1991) Five-year clinical study of u.v.-polymerized posterior composites. *Journal of Dentistry* **19**: 214-20
- Wilkie R, Lidums A and Smales RJ (1993) Class II glass ionomer cermet tunnel, resin sandwich and amalgam restorations over 2 years. *American Journal of Dentistry* **6**: 181-4

- Wilson MA, Wilson NH and Smith GA (1986) A clinical trial of a visible light-cured posterior composite resin restorative: two-year results. *Quintessence International* **17**: 151-5
- Wilson NH, Wastell DG and Norman RD (1996) Five-year performance of high-copper content amalgam restorations in a multiclinical trial of a posterior composite. *Journal of Dentistry* **24**: 203-10
- Wilson NH, Wilson MA, Offtall DG *et al* (1991) Performance of occlusin in butt-joint and bevel-edged preparations: five-year results. *Dental Materials*. **7**: 92-8
- Wilson NH, Wilson MA, Wastell DG *et al* (1988) A clinical trial of a visible light cured posterior composite resin restorative material: five-year results. *Quintessence International* **19**: 675-81
- Wilson NH, Wilson MA and Smith GA (1985) A clinical trial of a new visible light-cured composite restorative- initial findings and one-year results. *Quintessence International* **16**: 281-90
- Wilson NHF, Burke FJT and Mjör IA (1997) Age of restorations at replacement in permanent teeth in general dental practice *Acta Odontologica Scandinavica* **58**: 97-101
- Wilson NHF, Burke FJT and Mjör IA (1997) Reasons for placement and replacement of restorations of direct restorative materials by a selected group of practitioners in the United Kingdom. *Quintessence International* **28**: 245-8

- Winn DM, Brunelle JA, Selwitz RH *et al* (1996) Coronal and root caries in the dentition of adults in the United States, 1988-1991. *Journal of Dental Research* **75 (special issue)**: 642-651
- Worthington H, Holloway P, Clarkson J *et al* (1997) Predicting which adult patients will need treatment over the next year. *Community Dentistry and Oral Epidemiology* **25**: 273-7
- York AK and Arthur JS (1993) Reasons for placement and replacement of dental restorations in the United States Navy Dental Corps. *Operative Dentistry* **18**: 203-8
- York AK and Stephen AL (1993) Reasons for placement of dental restorations in the United States Navy dental corps. *Operative Dentistry* **18**: 203-8

APPENDIX

Appendix 1

The USPHS criteria

1a Caries assessment

This test is a visual inspection with mirror and probe if needed.

Code A: This is scored if there is no evidence of caries contiguous with the margin of the restoration

Code B: This is scored if there is evidence of caries contiguous with the margin of the restoration.

An area at the restoration margin is carious if a probe “catches” or resists removal after insertion with moderate to firm pressure and is accompanied by one or more of the following

- a. softness**
- b. opacity at the margin, as evidence of undermining or demineralisation**
- c. etching or a white spot as evidence of demineralisation**

An area at the margin is also considered carious if the probe does not “catch” but the conditions b or c are present.

1b Marginal adaptation

This test is a visual and physical test that involves the use of a mirror and probe which is lightly drawn a sharp probe across the restoration margin.

Code A: This is scored if there **is no** evidence of a crevice along the restoration margin which a probe penetrates.

Code B: This is scored if there **is** evidence of a crevice along the restoration margin which a probe penetrates but **no** dentine or base is exposed.

Code C: This is scored if there **is** evidence of a crevice along the restoration margin which a probe penetrates **and** dentine or base is exposed.

Code D: This is scored if there **is** evidence of a crevice along the restoration margin which a probe penetrates **and** dentine or base is exposed **and** the restoration is mobile, fractured or missing in part or *in toto*.

1c Marginal discolouration

This is a visual inspection.

Code A: This is scored if there **is no** evidence of any discolouration on the margin between the restoration and adjacent tooth structure.

Code B: This is scored if there **is** evidence of discolouration on the margin between the restoration and adjacent tooth structure but it does not involve more than 10% of the restoration's margin.

Code C: This is scored if there **is** evidence of discolouration on the margin between the restoration and adjacent tooth structure **and** it involves more than 10% of the restoration's margin.

1d Colour match

This is a visual test which utilises a dental mirror when necessary.

Code A: This is scored if there **is no** mismatch in colour, shade and / or translucency between the restoration and the adjacent tooth structure.

Code B: This is scored if there **is** a mismatch in colour, shade and / or translucency between the restoration and the adjacent tooth structure **but is** within the normal range of tooth colour, shade and / or translucency.

Code C: This is scored if there **is** a mismatch in colour, shade and / or translucency between the restoration and the adjacent tooth structure **and is** out with the normal range of tooth colour, shade and / or translucency.

1e Anatomical form

This is a visual test which uses a mirror if necessary.

Code A: This is scored if the restoration is continuous with the existing anatomical form of the tooth and there is no evidence the restoration is under-contoured.

Code B: This is scored if the restoration is continuous with the existing anatomical form of the tooth **but there is** evidence the restoration is under-contoured.

Code C: This is scored if the restoration is continuous with the existing anatomical form of the tooth, **there is** evidence the restoration is under-contoured **and** there is sufficient material missing to expose dentine or base.

Appendix 2

Examples of results sheets and graphic material used

Appendix 2.1 Result sheet for determining restorative status by BC and RMc i.e. determination of the "gold standard"

[illegible]

Appendix 2.2 Initial assessment sheet for assessors

Assessments and Recordings

Identification code: 67890....

Please examine the restoration(s) on the following surface(s) of the tooth indicated. Note your “feeling” with respect to leaving or replacing the restoration. If you feel that the restoration should be replaced then please tell your scribe your reason as to why so it can be recorded.

Model set 67890 RIGHT SIDE.....

Tooth	Surface	Replace? Yes or No	If yes, why?
First premolar	Disto-occlusal		
Second premolar	Disto-occlusal		
First molar	Occlusal		
Second molar	Occlusal		

Model set 67890 LEFT SIDE.....

Tooth	Surface	Replace? Yes or No	If yes, why?
First premolar	Disto-occlusal		
Second premolar	Disto-occlusal		
Second premolar	Mesio-occlusal		
First molar	Occlusal		
Second molar	Occlusal		

Appendix 5 Numbers of registered general dental practitioners in the UK as of 1/1/2009 (data supplied by the General Dental Council)

Age	Number of registered dentists
< 21	0
22-30	7102
31-40	10370
41-50	9605
51-60	6638
61-65	1600
> 65	861

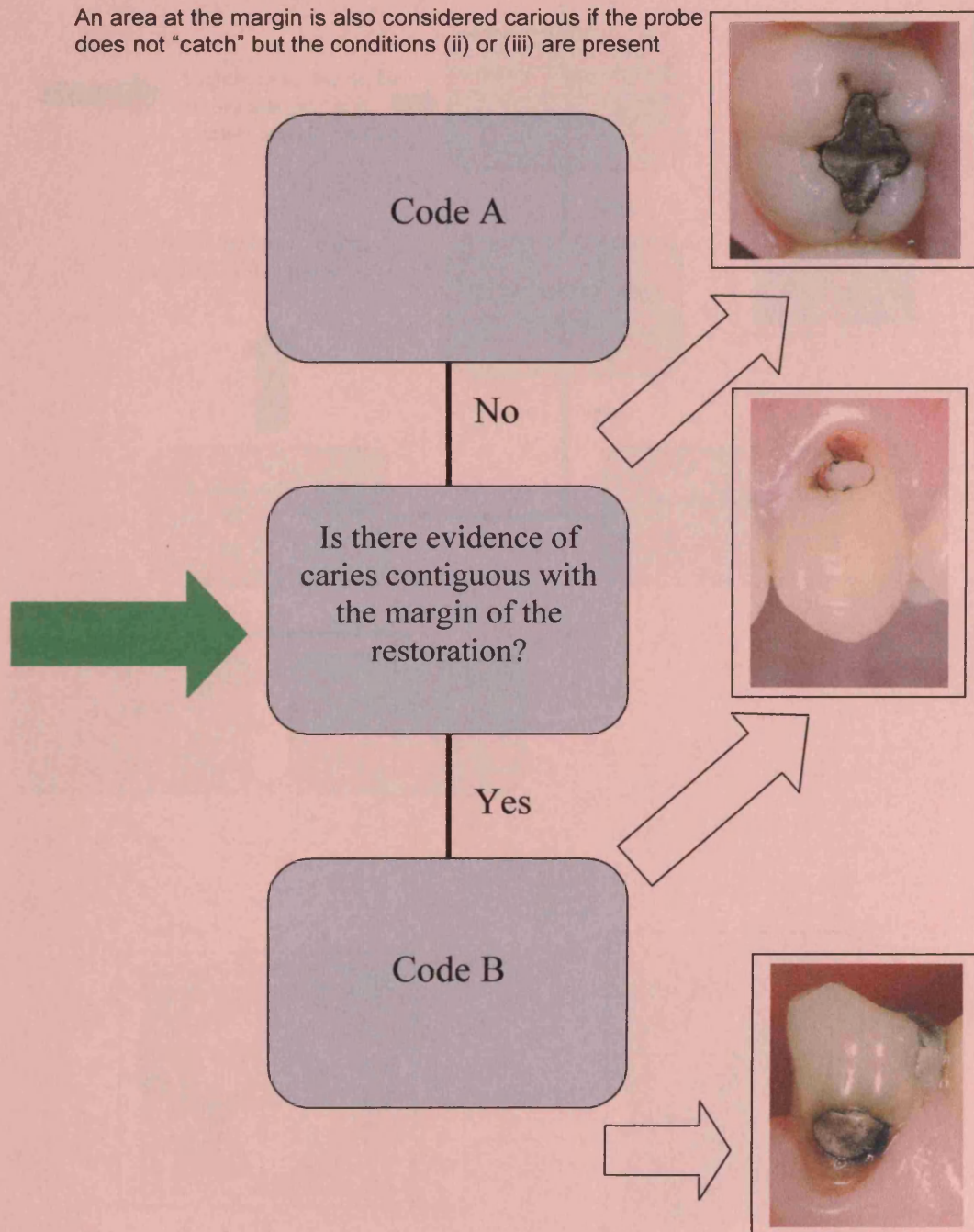
Appendix 2.3 a

Caries Assessment

An area at the restoration margin is carious if a probe “catches” or resists removal after insertion with firm to moderate pressure and is accompanied by one or more of the following.

- i. softness
- ii. opacity at the margin, as evidence of undermining or demineralisation
- iii. etching or a white spot as evidence of demineralisation

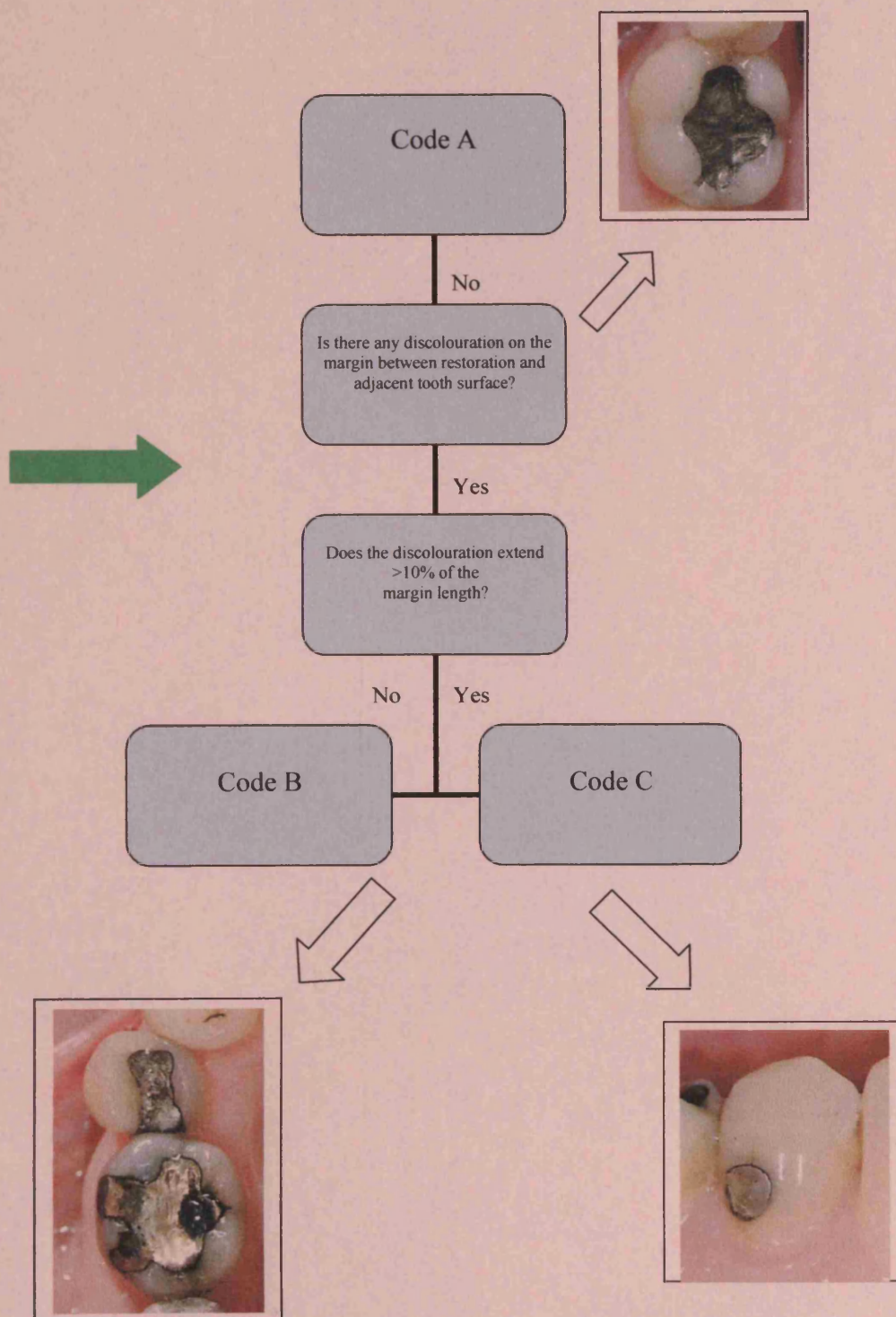
An area at the margin is also considered carious if the probe does not “catch” but the conditions (ii) or (iii) are present



Appendix 2.3c

Marginal Discolouration

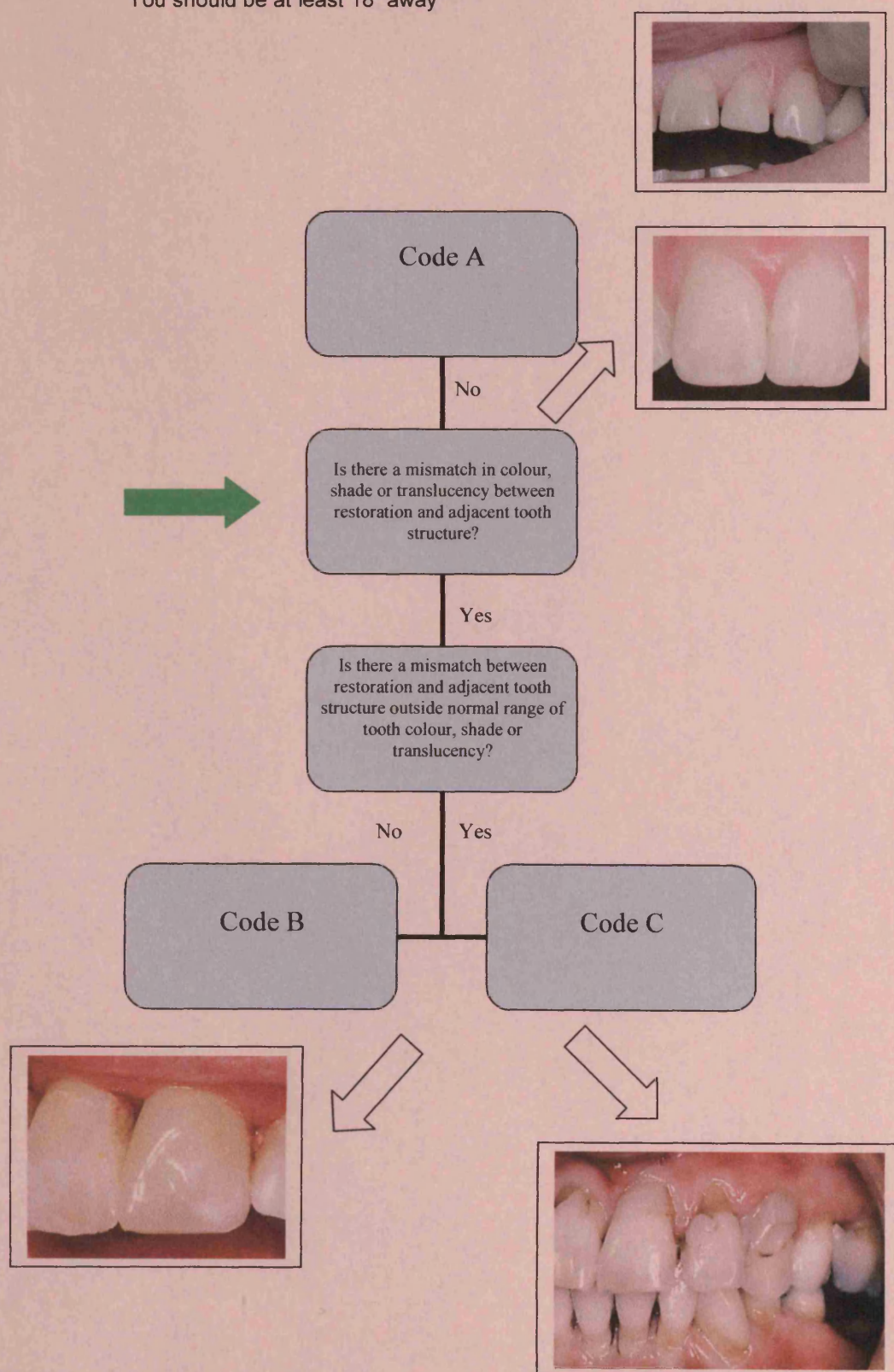
This is a visual test! Use a mirror if you need to.



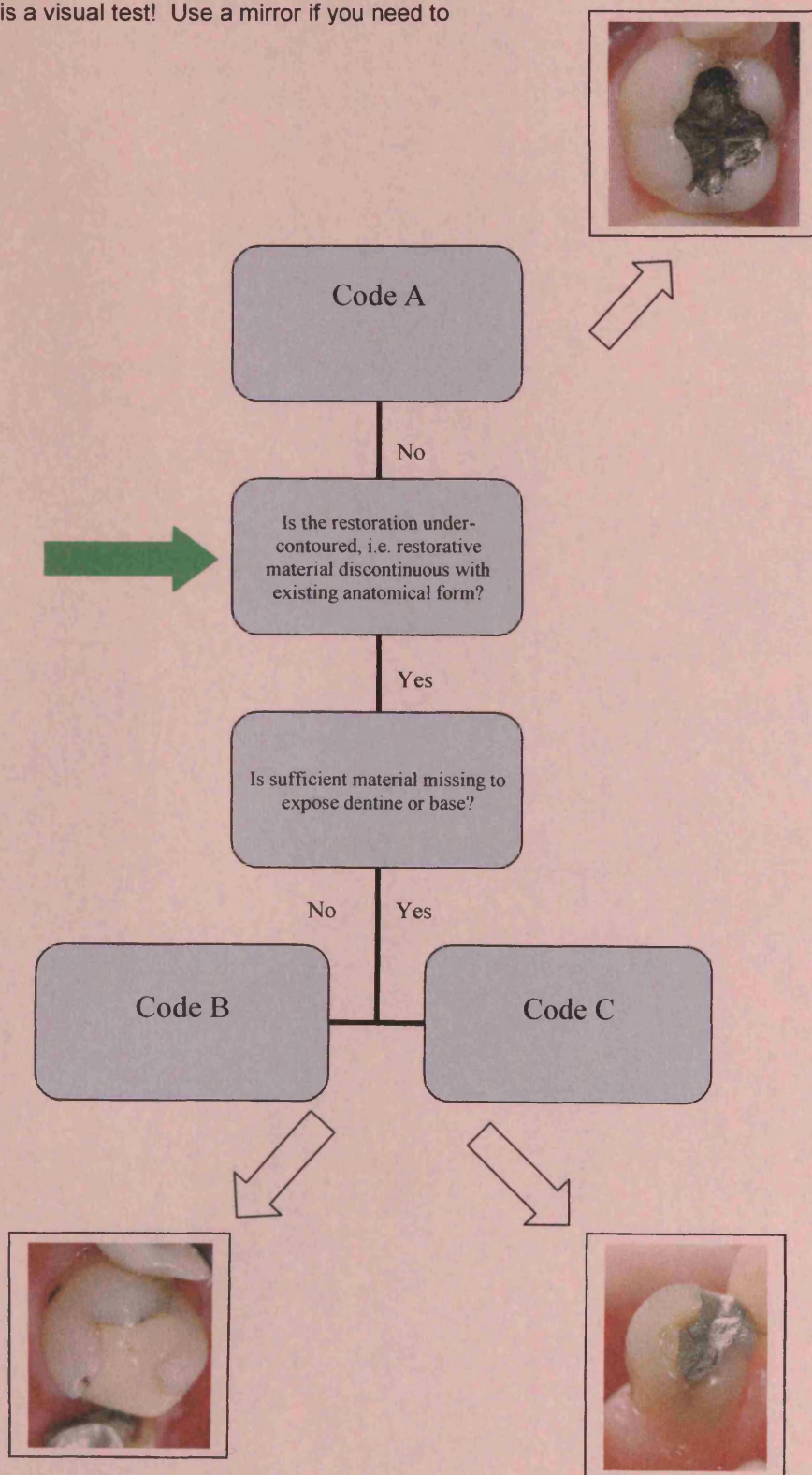
Appendix 2.3 d

Colour Match

This is a visual test! Use a mirror if you need to.
You should be at least 18" away



This is a visual test! Use a mirror if you need to



Appendix 2.4

Examples of photographs used in the training programme



Appendix 3.1 Copy of ethical approval



AWDURDOD IERCHYD
BRO TAF
HEALTH AUTHORITY

16th May 2001

AD/CP/JJL

Mr R McAndrew,
Senior Lecturer,
Adult Dental Health,
University Dental Hospital,
Heath Park,
Cardiff.

Dear Mr McAndrew,

01/3963 - Applying evaluation criteria to assess the need to replace dental restorations in Community Dental Practice - A pilot study looking at sensitivity, specificity and examiner variability

The Bro Taf Local Research Ethics Committee (Panel B) reviewed the above application for ethical approval at its meeting on the 16th May 2001. I am pleased to be able to inform you that ethical approval was granted subject to the following conditions:-

1. The Panel agreed that a more detailed patient information sheet and consent form should be provided. The Patient Information Sheet and Consent Form should also be provided on locally headed notepaper. Please note that it is also Bro Taf Local Research Ethics Committee policy to insist that all consent forms allow space for the both the signature of a witness and an investigator.

Your research may not proceed until you have complied with the conditions of approval. A formal written response is required indicating your compliance and attaching any amended or additional documentation to the Executive Officer of the Local Research Ethics Committee at the above address. I will consider your response and if satisfactory a letter will be sent to you indicating that your research may proceed.

You will no doubt realise that whilst the Local Research Ethics Committee has given approval for your project on ethical grounds, it is still necessary for you to obtain approval, if you have not already done so, from the relevant NHS Trust and /or College Office of Research & Development in which the work will be carried out.

I enclose for your information a copy of the Bro Taf Membership list on which the Members of Panel B, who were present at the meeting on the 16th May 2001, are indicated. I confirm that the Bro Taf Local Research Ethics Committee complies with the ICH Guidelines for Good Clinical Practice as they relate to an Independent Ethics Committee. A copy of the Committee's Constitution and Terms of Reference is available on request.



☐ HEADQUARTERS:
Churchill House
17 Churchill Way, Cardiff, CF10 2TW
PRIF SWYDDFA:
Tŷ Churchill
Ffordd Churchill, Caerdydd, CF10 2TW

Tel: 029 20 402402

☒ Temple of Peace and Health
Cathays Park, Cardiff, CF10 3NW

Tenil Hoddwch ac Iechyd:
Parc Cathays, Caerdydd, CF10 3NW

Fax/Ffacs: 029 20 402403

WHITN: 1809



Appendix 3.1b Copy of trust utilisation of resource approval



NHS
CYMRU
WALES

Eich cyf/Your ref
Ein cyf/Our ref
Welsh Health Telephone Network 1872
Direct line/Ulinell unlongrychol

Cardiff and Vale NHS Trust

Ymddiriedolaeth GIG
Caerdydd a'r Fro

**University Hospital of Wales
Ysbyty Athrofaol Cymru**

Heath Park,
Cardiff, CF14 4XW
Phone 029 2074 7747
Fax 029 2074 3838
Minicom 029 2074 3632

Parc Y Mynydd Bychan,
Caerdydd, CF14 4XW
Ffôn 029 2074 7747
Ffacs 029 2074 3838
Minicom 029 2074 3632

Tel: 029 20743742
Fax: 029 20745311
E-mail: evansmv@cf.ac.uk

From: Mary Evans
Trust R&D administrator
Radnor House
UHW

27 July, 2001

Mr R McAndrew
Adult Dental Health
Dental Hospital
UWCM

Dear Mr McAndrew,

Project Ref: 01/dh/1276

Applying evaluation criteria to assess the need to replace dental restorations in community dental practice - a pilot study looking at sensitivity, specificity and examiner variability

The above project has been received by the Trust R&D Office; it has been reviewed to determine any costs incurred by the Trust and I am pleased to confirm that the support costs of this research will be met by the Trust's R&D Support Funding allocation. As a result, the Trust is happy for you to continue with your application for ethical approval.

Please ensure that a copy of this letter accompanies your application to the Local Research Ethics Committee.

May I take this opportunity to wish you well with your study.

Yours sincerely,

Mary Evans
R&D Administrator
A:\july27-2.doc



Appendix 3.2a Consent sheet for volunteering to take part in project

Coleg Meddygaeth Prifysgol Cymru University of Wales College of Medicine



Dental School
Department of Adult Dental Health

Head of Department: Professor Paul Dummer

Direct ☎ +44 (029) 20742614
Direct FAX +44 (029) 20743120
Department ☎ +44 (029) 20744356
E-mail mcandrew@cardiff.ac.uk
RM/JO

1 October 2003

Mr K. Burford
Part-time Lecturer
Dept of Adult Dental Health

Dear Kevin

You may be aware that I am at present trying to recruit qualified dentists to take part in a "clinical trial". I aim to determine whether or not training in restoration evaluation has an effect on a decision to replace or leave a filling. The study will be partly laboratory based and partly clinical. The first part and laboratory phase would involve looking at some restorations and recording whether or not you feel they should be replaced.

I enquire as to whether or not you would feel able to help me with this part of my study. I would hope that it could be carried out at a mutually convenient time when you are at the Dental Hospital. If you are interested in taking part, please could you return the tear off slip attached to this letter.

With thanks.

Yours sincerely

Mr R McAndrew
BDS, MScD, FDS, FDS(Rest)RCS, DRD, MRDRCS(Ed)
Clinical Senior Lecturer and Honorary Consultant in Restorative Dentistry

X -----

NAME KEVIN BURFORD

Yes I would like to help if I can ☒

No I feel unable to help at the present time ☐



THE QUEEN'S
ANNIVERSARY PRIZE:
FOR HIGHER AND FURTHER EDUCATION
1998

Appendix 3.2b Consent form for taking part in phase 1

Dental School
Department of Adult Dental Health
Head of Department Professor Paul M H Dummer BDS MScD PhD DDSc FDSRCS(Ed)

Ysgol Ddeilintyddol
Adran Iecchyd Deintyddol i Oedollion
Pennaeth Adran Yr Athro Paul M H Dummer BDS MScD PhD DDSc FDSRCS(Ed)
Direct ☎ +44 (0)29 20742614
Direct FAX +44 (0)29 20743120
Department ☎ +44 (0)29 20744356
E-mail macandrew@cardiff.ac.uk

RM/JO

24 November 2004

Mrs S Oliver
Lecturer
Department of Adult Dental Health

Dear Sheila

I would like to thank you for agreeing to help with my research. I enclose a date where we can "potentially" get together. I would be grateful if you could let me know if any of the times and date are suitable. I will then confirm by e-mail.

Thursday 9th December - a.m.

Yours sincerely

Mr R McAndrew

Mr R McAndrew
BDS, MScD, FDS, FDS(Rest)RCS, DRD, MRDRCS(Ed)
Clinical Senior Lecturer and Honorary Consultant in Restorative Dentistry

X _____

Name: S. J. OLIVER

Date: 9/12/04.

Time: a.m. 9.15 - 10.45 11.00 - 12.30
(please circle)

Location: Oral Health Clinic *(Sorry I am on the Advance Schedule for the above session!)*

POSSIBLE ALTERNATIVE DATES: 7.12.04 AM
13.12.04 AM
14.12.04 AM
16.12.04 AM

Regards Sheila

REPRODUCTION OF THIS FORM
WITHOUT PERMISSION IS PROHIBITED

Professors / Athrawon
Prof P M H Dummer
Prof D H Edwards
Prof J S Rees

Readers
Darlennydd
Mr P H Jacobsen
Mr R G Jagger

Senior Lecturers / Uwch Ddarlithwyr
Dr A S M Gilmour
Dr S M Jenkins
Mr R McAndrew
Mr J Sweet
Dr S A Thompson
Mr J Wilson

Colleg meddygaeth
Ysgol meddygaeth
college of medicine



Cardiff University
Heath Park
Cardiff CF14 4XY, UK
Tel Ffôn +44(0)29 2074
Fax Ffôn +44(0)29 2074
E-mail E-bost adhdapt@cardiff.ac.uk
www.cardiff.ac.uk/dentistry
Prifysgol Caerdydd
Mynydd Bychan
Caerdydd CF14 4XY, DU

Appendix 3.2c Consent form for volunteer patients



Eich cyf/Your ref
Ein cyf/Our ref
Welsh Health Telephone Network 1872
Direct line/Llinell uniongyrchol

NHS
WALES
GIG
CYMRU

Cardiff and Vale NHS Trust

Ymddiriedolaeth GIG
Caerdydd a'r Fro

University Dental Hospital
Ysbyty Deintyddol Athrofaol

Heath Park
Cardiff
CF14 4XY
Phone 029 2074 7747

Parc Y Mynydd Bychan
Caerdydd
CF14 4XY
Ffôn 029 2074 7747

Applying evaluation criteria to assess the need to replace restorations in dental practice

Patient consent form

I
Of
.....
.....
.....
.....

hereby consent to take part in the above study.

- I have read the information sheet.
- I understand that the study may require me to attend the dental hospital on four separate occasions.
- I understand I can withdraw from the study at any time without prejudice
- The procedure involved have been explained to me by Mr Robert McAndrew

Signed
Subject

Signed
Clinician



Appendix 3.2d Patient information sheet



Eich cyf/Your ref
Ein cyf/Our ref
Welsh Health Telephone Network 1872
Direct line/Llinell uniongyrchol

NHS
WALES
GIG
CYMRU

Cardiff and Vale NHS Trust

Ymddiriedolaeth GIG
Caerdydd a'r Fro

University Dental Hospital Ysbyty Deintyddol Athrofaol

Heath Park
Cardiff
CF14 4XY
Phone 029 2074 7747

Parc Y Mynydd Bychan
Caerdydd
CF14 4XY
Ffôn 029 2074 7747

Department of Adult Dental Health Patient information sheet

Applying evaluation criteria to assess the need to replace restorations in dental practice

Thank you for taking time to read this information sheet and for considering taking part in this project. It goes without saying that without volunteers such as yourself clinical research would be extremely difficult if not impossible.

Background information

Dentists can disagree whether or not a particular filling needs to be replaced. This study is looking at whether or not training dentists to look at fillings in a particular way reduces the level of disagreement.

What is involved?

If you agree to take part you in the study you will need to visit the Dental Hospital on three or four separate occasions. You will have your teeth looked at by qualified dentists when you attend and the process will be very similar to that of a routine dental check-up. Each visit will last around two hours. The process should not be uncomfortable.

Do I have to take part?

No. Participation in the study is voluntary and you can withdraw at any time without it affecting the way you are treated at the Dental Hospital.

Will I be paid for taking part in the study?

No.

Who should I speak to if I have any questions about the project?

If you would like any more details about the project then please do not hesitate to ask Mr Robert McAndrew who is the lead researcher on the project.

Robert McAndrew
Clinical Senior Lecturer in
Restorative Dentistry

Tel: 02920 742614
E mail mcandrew@cf.ac.uk

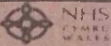


INVESTOR IN PEOPLE



Appendix 3.3

Photography consent form

 MEDICAL PHOTOGRAPHY REQUEST FORM Dental Illustration Unit Ext. 2606	
PATIENT'S DETAILS: Name _____ Address _____ DOB _____ Hospital Number _____ Patient label ESSENTIAL	CONSENT LEVELS (requesting clinician to explain to patient): 1. I consent to photographs/video (teaching(s)) being taken for my personal medical case-notes only. 2. I consent to photographs/video recordings being taken for my medical case-notes and being used for teaching of healthcare staff and students. <i>The patient has the right to withdraw their consent at any time by contacting the Dental Illustration Unit at University Dental Hospital.</i> 3. I consent to my photographs/video recordings being published in an open access journal, textbook or other form of medical publication (which may include the internet) and therefore may be seen by the general public as well as healthcare professionals. <i>Although the patient has the right to withdraw consent it is not possible to withdraw published material.</i>
DIAGNOSIS: PHOTOGRAPHS REQUIRED: (Please refer to the set views listed in the photographic protocols booklet on clinic) ☐ Standard Orthodontic ☐ Case Presentation ☐ Periodontic/Paediatric ☐ Gingival Other _____ This section to be completed by DU staff Of No. ☐ ☐ ☐ ☐ ☐ ☐ File No. _____ Photograph Date _____	REQUESTING CLINICIAN: Student No. ☐ ☐ ☐ Name (PLEASE PRINT) _____ Position _____ Signature _____ Date _____ Clinic _____ Consultant _____ CONSENT: ☐ 1 Case-notes (see above) Signature _____ Date _____ ☐ 2 Teaching (see above) Signature _____ Date _____ ☐ 3 Publication (see above) Signature _____ Date _____
DIAGNOSIS: PHOTOGRAPHS REQUIRED: (Please refer to the set views listed in the photographic protocols booklet on clinic) ☐ Standard Orthodontic ☐ Case Presentation ☐ Periodontic/Paediatric ☐ Gingival Other _____ This section to be completed by DU staff Of No. ☐ ☐ ☐ ☐ ☐ ☐ File No. _____ Photograph Date _____	REQUESTING CLINICIAN: Student No. ☐ ☐ ☐ Name (PLEASE PRINT) _____ Position _____ Signature _____ Date _____ Clinic _____ Consultant _____ CONSENT: ☐ 1 Case-notes (see above) Signature _____ Date _____ ☐ 2 Teaching (see above) Signature _____ Date _____ ☐ 3 Publication (see above) Signature _____ Date _____
DIAGNOSIS: PHOTOGRAPHS REQUIRED: (Please refer to the set views listed in the photographic protocols booklet on clinic) ☐ Standard Orthodontic ☐ Case Presentation ☐ Periodontic/Paediatric ☐ Gingival Other _____ This section to be completed by DU staff Of No. ☐ ☐ ☐ ☐ ☐ ☐ File No. _____ Photograph Date _____	REQUESTING CLINICIAN: Student No. ☐ ☐ ☐ Name (PLEASE PRINT) _____ Position _____ Signature _____ Date _____ Clinic _____ Consultant _____ CONSENT: ☐ 1 Case-notes (see above) Signature _____ Date _____ ☐ 2 Teaching (see above) Signature _____ Date _____ ☐ 3 Publication (see above) Signature _____ Date _____

Appendix 4.1 An explanation of calculating Dices Coincidence Index

RJUR6	2	2	2	2	1	2	2	2	2	2	2	2
RJUL6	2	1	2	2	2	2	2	2	2	2	2	2
RJLL6	2	1	2	2	2	2	2	2	2	2	2	2
JHUR7D	2	2	2	1	2	2	2	2	2	2	2	2
JHUR7M	2	2	2	2	2	2	2	2	2	2	2	2
JHUR6D	2	2	2	2	1	2	2	2	2	2	2	2
JHUR6M	2	2	2	2	2	2	2	2	2	2	2	2
JHUL6M	2	2	2	2	2	2	2	2	2	2	2	2
JHUL6D	2	2	2	2	2	2	2	2	2	2	2	2
JHUL7M	2	2	2	2	2	2	2	2	2	2	2	2
JHUL7D	2	2	2	2	2	2	2	2	2	2	2	2
JHLL7	2	2	2	2	2	2	2	2	2	2	2	2
JHLL6	2	2	2	2	1	2	2	2	2	2	2	2
JHIR6	2	2	1	2	1	2	2	2	2	2	2	2
JHIR7	2	2	2	2	2	2	2	2	2	2	2	2
JOUR6	2	1	2	2	2	2	2	2	2	2	2	2
JOUR5	2	1	1	2	1	2	1	2	1	2	1	2
JOUL4	2	2	2	2	2	2	2	2	2	2	2	2
JOUL5	2	2	2	2	2	2	2	2	2	2	2	2
JOUL6	2	2	1	2	2	2	2	2	2	2	2	2
JOUL7	1	1	2	2	2	2	2	2	2	2	2	2
JOLL7	2	2	2	2	2	2	2	2	2	2	2	2
JOLL5	2	2	2	2	2	2	2	2	2	2	2	2
JOLR5	2	2	2	2	2	2	2	2	2	2	2	2
JOLR6	2	1	2	2	2	2	2	2	2	2	2	2
JOLR7	2	2	2	2	1	2	2	2	2	2	2	2
CKUR7	2	2	1	2	1	2	2	2	2	2	2	2
CKUL5	2	2	2	2	2	2	2	2	2	2	2	2
CKUL6	2	2	2	2	2	2	2	2	2	2	2	2
CKLL7	2	2	2	2	2	1	2	2	2	2	2	2
CKLL6	2	2	2	2	2	2	2	2	2	2	2	2
CKLR6	2	2	2	2	2	2	2	2	2	2	2	2
CKLR7	2	2	2	2	2	2	2	2	2	2	2	2
RKUR7	2	1	2	1	1	1	2	1	2	1	2	1
RKUR6	2	2	2	2	2	2	2	2	2	2	2	2
RKUL6	1	1	1	2	1	1	2	1	1	1	1	1
RKUL7	2	2	2	2	2	2	2	2	2	2	2	2
RKLL7	2	1	2	2	2	1	2	2	2	2	2	2
RKLL6	2	2	2	2	2	2	2	2	2	2	2	2
RKLR6	2	2	2	2	2	1	2	2	2	2	2	2
RKLR7	2	2	2	2	1	2	2	2	2	2	2	2
SSUR6	2	2	2	2	2	2	2	2	2	2	2	2
SSUR5	2	2	2	2	2	2	2	2	2	2	2	2

The gold standard (column 2) dictates how the assessor has performed (columns onwards). In this instance there are 4 possible outcomes that can be calculated for each assessor:

- the gold standard and the assessor agree that a restoration is sound and does not need replaced (1x1) = True positive
- that the gold standard suggests a restoration is sound but the assessor disagrees (1x2) = False negative
- that the gold standard says that a restoration needs replaced but the assessor disagrees (2x1) = False positive
- the gold standard and assessor agree a restoration does not need to be replaced (2x2) = True negative

The score of the assessor compared to that of the gold standard can then be used to calculate the Dices' Coincidence index for agreement that a restoration needs replaced or not. The formula is detailed in Table 3.4 (p145)

Appendix 5 Numbers of registered general dental practitioners in the UK as of 1/1/2009 (data supplied by the General Dental Council)

Age	Number of registered dentists
< 21	0
22-30	7102
31-40	10370
41-50	9605
51-60	6638
61-65	1600
> 65	861