IMPROVING SURGICAL LEARNING AND PERFORMANCE AT UNICOMPARTMENTAL KNEE ARTHROPLASTY

A DISSERTATION SUBMITTED TO THE UNIVERSITY OF OXFORD



IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

ABTIN ALVAND

WOLFSON COLLEGE

TRINITY TERM 2014

ABSTRACT

Improving Surgical Learning and Performance at Unicompartmental Knee Arthroplasty

Abtin Alvand **Doctor of Philosophy**
Wolfson College **Trinity Term 2014**

Surgery is a difficult skill to teach and even more difficult to assess. In the context of orthopaedic surgery, the Oxford unicompartmental knee arthroplasty (OUKA) is a commonly performed and technically challenging procedure with a variable and sometimes unsatisfactory revision rate. Thus there is a need to develop methods for optimising performance in this procedure. This thesis aims to use this procedure to explore ways of improving teaching methods and assessment methods, and the use of novel technologies to impact surgical performance.

No validated technical skill assessment system currently exists to evaluate surgical performance during OUKA. A consensus-based framework of expert opinion (Delphi methodology) was therefore used to determine the essential technical steps of OUKA to aid development of a technical skill assessment system. A procedure-specific assessment scale named the Oxford Arthroplasty Technical Skill Score (OATSS) was produced. Two clinical studies were conducted to demonstrate the inter-rater reliability and construct validity of OATSS in the real operating theatre. This is the first time that a procedure-specific assessment scale has been developed and validated for a complex orthopaedic procedure.

The Delphi methodology was also used to identify important radiographic parameters that could be used to predict clinical outcome following OUKA. Two cohort studies used post-operative radiographs of "failed" and "poorly functioning" OUKA cases (with matched controls) to identify several radiographic parameters that can predict the increased risk of a poor clinical outcome.

A virtual reality (VR) based simulator (called ViTAL) was developed to teach the cognitive skills necessary for performing the OUKA procedure. This software-based learning environment was used in a randomised controlled study to assess its training effectiveness. This demonstrated that, when used alongside existing teaching modalities, ViTAL resulted in better acquisition of cognitive skills in novice trainees.

A randomised controlled study was subsequently undertaken using anatomic dry-model bones in a surgical skills laboratory to determine the effect of cognitive skills training using the ViTAL simulator on the ability of orthopaedic trainees to perform the OUKA procedure. The previously validated OATSS assessment scale was used to evaluate surgical performance. Results demonstrated that this type of training not only improved acquisition of cognitive skills, but also resulted in better technical skill performance during OUKA procedures. This demonstrates, for the first time, the effectiveness of a purely cognitive VR-based simulator for improving technical skills during arthroplasty surgery.

Patient-specific instrumentation (PSI) is a novel and exciting technological development within orthopaedic surgery. A prospective RCT was performed to compare OUKAs implanted using PSI with those implanted using conventional instrumentation. Intra-operative measurements, together with post-operative radiographic analyses, were performed to determine the accuracy of component positioning. With regards to positioning of the femoral component, there was no difference between the accuracy of PSI and conventional instrumentation. However, accuracy of PSI was poor with regards to positioning of the tibial component. Further improvement to the PSI is needed before this can be recommended for routine use.

In conclusion, this thesis has demonstrated that improvement of learning and assessment methods is currently the safest and most feasible means of optimising surgical performance during the OUKA procedure. Although it is important to embrace novel training modalities and technological developments in an attempt to enhance surgical performance, they must be scientifically evaluated to ensure that they deliver real clinical improvements.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my three supervisors, Prof. Jonathan Rees, Prof. David Murray and Prof. Richie Gill. They have steadfastly guided me through this journey and I am grateful for their inspiration, teaching and support.

A special thank you must go to Mrs Barbara Marks and Mrs Jo Brown for their invaluable help with administrative support and organisation of several projects.

I am very grateful to Prof. Andrew Price and Prof. David Beard for their valuable academic advice, ideas and interest in my thesis. I would also like to thank Mr Chris Dodd and Mr Will Jackson for their help and expertise during the clinical studies.

Dr Minsi Chen's computer science expertise was instrumental in the development of the ViTAL simulator and I am very grateful for all his help throughout this project. I would also like to thank Mr Max Biege for his graphic design expertise.

Completing this Thesis would have been all the more challenging if it wasn't for the support and friendship of the "Boys": Tanvir, George, Chethan, Steve, Ben K, Sami, Raj, Ben D, Luke, Paul, Geraint and Alex.

I will forever be indebted to my parents, Daryoush and Fataneh, and sister Tina for their encouragement, help and generosity throughout my academic career and life in general. The biggest thanks must go to my wife Vanessa for her never-ending love, patience and self-sacrifice. I dedicate this Thesis to my beautiful children Ava, Zara and Maani.

GLOSSARY

ACGME	Accreditation Council for Graduate Medical Education
AP	Anteroposterior
ASA	American Society of Anaesthesiologists
BMI	Body Mass Index
CT	Computer tomography
CI	Contextual interference
DOPS	Direct Observation of Procedural Skill
GRS	Global Rating Scale
ISCP	Intercollegiate Surgical Curriculum Project
ICC	Intraclass correlation coefficient
IQR	Interquartile range
ISTC	Independent Sector Treatment Centres
LA	Lateral (radiograph)
MMC	Modernising Medical Careers
NJR	National Joint Registry
NOC	Nuffield Orthopaedic Centre
OA	Osteoarthritis
OATSS	Oxford Arthroplasty Technical Skill Score
OCAP	Orthopaedic Competency Assessment Project

OKS	Oxford knee score
OSATS	Objective Structured Assessment of Technical Skill
OUKA	Oxford unicompartmental knee arthroplasty
PBA	Procedure-Based Assessments
PSI	Patient-specific instrumentation
RCT	Randomised controlled trial
SD	Standard deviation
TKA	Total knee arthroplasty
UKA	Unicompartmental knee arthroplasty
VR	Virtual Reality
ViTAL	Virtual Reality Tool for Arthroplasty Learning
3D	Three-dimensional

DEFINITION OF COMMON TERMS

Internal consistency: A measure of reliability that ensures different test items measuring a construct deliver consistent scores. With regards to surgical skill assessment, the assumption is that a skilled surgeon will rate highly (although not identically) throughout all the different domains of a surgical skill assessment system.

Reliability: A measure of the reproducibility of test scores obtained from an assessment given multiple times under the same conditions. (*Interobserver* reliability measures the agreement between different assessors for a given test, whereas *intraobserver* reliability determines the variability of a single assessor's marks for the same test on different occasions)

Usability: The effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments (for example whilst using a virtual reality simulator for learning a surgical procedure).

Validity (Concurrent): The extent to which the results of an assessment tool correlate with the "gold" standard tests known to measure the same skill domain. High correlation between different assessment tools indicates good concurrent validity.

Validity (Construct): The most relevant type of validity when evaluating a technique for surgical skill assessment and refers to its ability to distinguish between subjects with varying levels of surgical expertise.

Validity (Content): Established by showing that the components of a test are relevant and represent the skills the evaluator is interested in. It ensures that the steps within a task are thought out and linked.

Validity (Face): Subjective measure of how closely an assessment tool or task resemble the situation in real life and is usually measured through questioning experts.

Validity (Predictive): The extent to which future performance can be predicted by the test score of an assessment tool.

TABLE OF CONTENTS

1. GENERAL INTRODUCTION AND LITERATURE REVIEW 1
1.1. INTRODUCTION 1
1.2. SURGICAL COMPETENCE 2
1.3. SURGICAL EXPERTISE 3
1.3.1. Automaticity 4
1.3.2. Error detection ability and diagnostic ability 5
1.3.3. The influence of surgical volume 6
1.3.4. Deliberate practice 6
1.4. What kind of practice makes perfect? 7
1.4.1. Distributed practice and the spacing effect 7
1.4.1.1. Contextual interference 8
1.5. LEARNING IN THE DOMAIN OF SURGERY 9
1.5.1. Rasmusson's model of human behaviour during complex tasks.....9
1.5.1.1. Learning psychomotor skills 10
1.5.1.2. The surgical learning curve and innate ability 12
1.6. SURGICAL EDUCATION: PAST AND PRESENT 13
1.6.1. Surgical apprenticeship and the Halstedian era 13
1.6.2. Recent challenges facing the delivery of surgical education.....13
1.7. ASSESSMENT OF SURGICAL SKILLS 15
1.7.1. What are the basic requirements of assessment tools? 15
1.7.2. Validity 15
1.7.3. Reliability 16
1.7.4. Current methods of assessing surgical skills 17
1.7.4.1. Observational tools for skill assessment.....17

1.7.4.2. Motion analysis and measurement of force patterns 18

1.8. ERRORS AND ADVERSE EVENTS IN SURGERY 20

1.9. RECENT METHODS EMPLOYED FOR IMPROVING SURGICAL EDUCATION 21

1.9.1. Computer-aided learning and web-based educational resources 21

1.9.2. Simulation-based surgical education 21

1.10. KNEE OSTEOARTHRITIS AND UNICOMPARTMENTAL KNEE ARTHROPLASTY 25

1.10.1. The burden of knee osteoarthritis 25

1.10.2. Unicompartmental knee arthroplasty 29

1.10.3. Oxford unicompartmental knee arthroplasty (OUKA) 30

1.10.4. Current methods of teaching and improving performance during the OUKA procedure 33

1.10.5. Chapter summary and rationale of this thesis 33

1.11. SPECIFIC AIMS 34

1.12. THESIS STRUCTURE 34

2. DEVELOPMENT AND VALIDATION OF A TECHNICAL SKILL ASSESSMENT TOOL FOR UNICOMPARTMENTAL KNEE ARTHROPLASTY 36

2.1. INTRODUCTION 36

2.1.1. Procedure-specific assessment scales 36

2.1.2. Use of a consensus-based Delphi approach 37

2.1.3. Reliability and validity 38

2.1.4. Aims 38

2.2. STUDY A: CONSENSUS-BASED FRAMEWORK FOR DEVELOPMENT OF THE “OXFORD ARTHROPLASTY TECHNICAL SKILL SCORE” (OATSS) 39

2.2.1. Methods 39

2.2.1.1. Study design and participant selection 39

2.2.1.2. Delphi design and determination of group consensus 40

2.2.1.3. Creation of the final technical skill evaluation tool for OUKA 42

2.2.1.4. Statistical analysis 42

2.2.2. Results 42

2.3. STUDY B: DETERMINING THE INTER-RATER RELIABILITY OF THE OXFORD ARTHROPLASTY TECHNICAL SKILL SCORE (OATSS) 45

2.3.1. Methods 45

2.3.1.1. Setting and participants 45

2.3.1.2. Assessment of OUKA performance 45

2.3.1.3. Statistical analysis 46

2.3.2. Results 46

2.4. STUDY C: DETERMINING THE CONSTRUCT VALIDITY OF THE OXFORD ARTHROPLASTY TECHNICAL SKILL SCORE (OATSS).....48

2.4.1. Methods 48

2.4.1.1. Setting and participants 48

2.4.1.2. Assessment of OUKA performance 48

2.4.1.3. Statistical analysis 49

2.4.2. Results 49

2.5. DISCUSSION.....50

2.5.1. Implications of the studies.....50

2.5.2. Limitations.....55

2.5.3. Summary.....56

3. DEVELOPMENT AND EVALUATION OF A POST-OPERATIVE RADIOLOGICAL ASSESSMENT SYSTEM FOR UNICOMPART-MENTAL KNEE ARTHROPLASTY 57

3.1. INTRODUCTION.....57

3.1.1. Aims 59

3.2. STUDY A: Consensus-based Delphi Approach for Determining Important Parameters for Analysis During Post-Operative Radiological Asssessment of OUKA.....59

3.2.1. Methods 59
3.2.1.1. Study design and participant selection 59
3.2.1.2. Delphi design and determination of group consensus 59
3.2.1.3. Selection of radiographic parameters for assessment of OUKA 61
3.2.1.4. Statistical analysis 62
3.2.2. Results 62
3.3. STUDY B: Determining the value of post-operative radiographic parameters in predicting clinical outcome following OUKA 69
3.3.1. Methods 69
3.3.1.1. Setting and patient cohorts 69
3.3.1.2. Radiological assessment 72
3.3.1.3. Statistical analysis 73
3.3.2. Results 73
3.4. DISCUSSION 79
3.4.1. Implications of the studies 79
3.4.2. Limitations 82
3.4.3. Conclusions 82
4. DEVELOPMENT AND VALIDATION OF A COGNITIVE SIMULATOR FOR TEACHING THE OXFORD UNICOMPARTMENTAL KNEE ARTHROPLASTY PROCEDURE 84
4.1. INTRODUCTION 84
4.1.1. The role of non-technical skills in surgery 85
4.1.2. Are non-technical skills important in UKA surgery? 85
4.1.3. Aims: 86
4.2. METHODS 86
4.2.1. Development of the cognitive simulator 86
4.2.1.1. Determining the content of the ViTAL simulator 90
4.2.2. Usability study to improve ViTAL 90

4.2.3. Randomised controlled trial to determine the effectiveness of ViTAL 93

4.2.3.1. Training protocol 94

4.2.3.2. Assessment of cognitive skills 96

4.2.3.3. Statistical analysis 96

4.3. RESULTS 98

4.3.1. Difficulty in teaching the surgical steps of the OUKA procedure 98

4.3.2. Usability study 100

4.3.3. Determining the effectiveness of ViTAL 102

4.4. DISCUSSION 105

4.4.1. Implications of the study 105

4.4.2. Limitations 109

4.4.3. Summary 110

5. RANDOMISED CONTROLLED TRIAL TO DETERMINE THE IMPACT OF COGNITIVE SKILLS TRAINING USING THE ViTAL SIMULATOR ON PERFORMANCE DURING SIMULATED UNICOMPARTMENTAL KNEE ARTHROPLASTY 111

5.1. INTRODUCTION 111

5.1.1. Aims: 112

5.2. METHODS 113

5.2.1. Setting and participants 113

5.2.1.1. Cognitive skills training 114

5.2.1.2. Assessment of cognitive skills 115

5.2.1.3. Post-training assessment of technical skills 116

5.2.2. Outcomes and statistical analysis 121

5.3. RESULTS 123

5.3.1. Cohort demograpchics 123

5.3.2. Determining the effectiveness of ViTAL simluator training for improving cognitive skills 124

5.3.3. Determining the effectiveness of ViTAL simluator training for improving technical skills....125

5.4. DISCUSSION....126

5.4.1. Implications of the study....126

5.4.2. Limitations....130

5.4.3. Summary....131

6. RANDOMISED CONTROLLED TRIAL TO ASSESS THE IMPACT OF PATIENT-SPECIFIC INSTRUMENTATION ON SURGICAL ACCURACY DURING THE OXFORD UNICOMPARTMENTAL KNEE ARTHROPLASTY PROCEDURE....132

6.1. INTRODUCTION....132

6.1.1. Aim:....135

6.2. METHODS....135

6.2.1. Study design and patients....135

6.2.2. Outcome measures....141

6.2.2.1. Primary outcome measures....141

6.2.2.1.1. Intra-operative assessment of surgical accuracy:....141

6.2.2.1.2. Post-operative radiological assessment of component positioning: 142

6.2.2.2. Secondary outcome measures....144

6.2.3. Statistical analysis....144

6.3. RESULTS....146

6.3.1. Participant demographics....146

6.3.2. Intra-operative assessment of surgical accuracy....146

6.3.2.1. Requirement for a tibial plateau “re-cut”:....146

6.3.2.2. Bearing size:....146

6.3.2.3. Bearing tracking:....147

6.3.2.4. Comparison of the implanted and planned component sizes:....148

6.3.2.5. Thickness of resected bone:....148

6.3.3. Radiological assessment of component positioning.....150
6.3.4. Secondary outcome measures.....152
6.3.5. Surgeons' overall impression of the PSI system.....153
6.4. DISCUSSION.....154
6.4.1. Implications of the study.....154
6.4.2. Limitations.....159
6.4.3. Summary.....160
7. DISCUSSION.....161
7.1. SUMMARY OF MAIN FINDINGS AND CONCLUSIONS.....161
7.1.1. Contributions to the field of study.....165
7.2. CLINICAL RELEVANCE.....167
7.3. RECOMMENDATIONS FOR FUTURE WORK.....169
APPENDICES.....172
Appendix 1: Comparison of radiographic parameters in Cohort 1.....172
Appendix 2: Comparison of radiographic parameters in Cohort 2.....173
Appendix 3: Comparison of of radiographic parameters in Cohort 3.....174
Appendix 4: ViTAL Usability Study: Participant demographics.....175
Appendix 5: Genre of video games played by participants.....177
Appendix 6: MCQs used for evaluation of the ViTAL Cognitive simulator.....178
REFERENCES:.....183

TABLE OF FIGURES

Figure 1-1 Motion analysis system being used during a simulated arthroscopy task on a phantom knee model. 19

Figure 1-2 Anatomical diagram of the knee joint (Patella hasbeen excluded to demonstrate intra-articular structures). 26

Figure 1-3 Anteroposterior radiograph of a knee demonstrating medial compartment osteoarthritis. 27

Figure 1-4 Excised articular surface of the medial tibial plateau demonstrating the classic wear pattern of anetromedial osteoarthritis. 28

Figure 1-5 Post-operative radiographs comparing total knee arthroplasty (left) and unicompartmental knee arthroplasty (right). 30

Figure 1-6 The Oxford unicompartmental knee arthroplasty. (Image courtesy of Biomet Inc) 31

Figure 1-7 Intraoperative photograph of minimally invasive Oxford unicompartmental knee arthroplasty, demonstrating the limited view of the knee joint through the small incision that adds to the technical difficulty of the procedure. (Image courtesy of Biomet Inc). 32

Figure 2-1 Delphi survey to identify: i) key steps in performing Oxford Unicompartmental Knee Arthroplasty; ii) whether or not these steps are easy to teach. 41

Figure 2-2. The Oxford Arthroplasty Technical Skill Score (OATSS) evaluation form. 44

Figure 2-3 Box plots comparing the performance of the Expert (n=13) and Trainee (n=13) cohorts during Oxford unicompartmental knee arthroplasty. (OATSS = Oxford Arthroplasty Technical Skill Score. Maximum possible OATSS = 90. Box represents the interquartile range, bold line in the box represents the median, whiskers represent the 2.5 and 97.5 percentiles, and circles are outliers) 50

Figure 3-1 Delphi survey to identify important parameters for analysis during post-operative radiographic assessment of OUKA. 61

Figure 3-2 Data collection form composed of important radiographic parameters used during post-operative radiological assessment of OUKA. 64

Figure 3-3 Post-operative AP and Lateral radiographs for assessment of component alignment, expressed as positive or negative degrees (A = Femoral component coronal alignment; B = Tibial component coronal alignment; C = Femoral component sagittal alignment; D = Tibial component posterior tilt).....66

Figure 3-4 AP post-operative radiograph following medial cemented OUKA highlighting important parameters. 67

Figure 3-5 Lateral post-operative radiograph following medial cemented OUKA highlighting important parameters. 68

Figure 4-1: Screenshot of the ViTAL cognitive simulator user-interface demonstrating the “Home Screen”. 89

Figure 4-2: Screen shot of the ViTAL cognitive simulator user-interface demonstrating the “femoral drill guide assembly”. 89

Figure 4-3. Questionnaire used in the usability study to improve the ViTAL cognitive simulator. 92

Figure 4-4: Flow diagram of participants through trial. 94

Figure 4-5: Performance of the medical students in the pre- and post-training MCQ tests for assessment of the cognitive skills necessary to perform the OUKA procedure (box represents the interquartile range, bold line in the box represents the median, whiskers represent the 2.5 and 97.5 percentiles, and circles are outliers). 4.5.....104

Figure 5-1: Flow diagram of the participants (orthopaedic trainees) through the trial.....114

Figure 5-2: Set-up of OUKA instrumentation, knee simulator, and webcams for evaluation of technical skills after training. 118

Figure 5-3: The Oxford Arthroplasty Technical Skill Score evaluation form used by the expert evaluators to assess the performance of participants during the simulated OUKA procedure. Scores for each of the 19 items were summed to generate the total OATSS value (maximum score: 90).....120

Figure 5-4: Performance of the orthopaedic trainees in the pre- and post-training MCQ tests for assessment of the cognitive skills necessary to perform the OUKA procedure (box represents the interquartile range, bold line in the box represents the median, whiskers represent the 2.5 and 97.5 percentiles, and the circles are outliers).....124

Figure 5-5: Post-training assessment of OUKA performance using the Oxford Arthroplasty Technical Skill Scale (OATSS) scoring system (maximum possible

score = 90, box represents the interquartile range, bold line in the box represents the median, and the whiskers represent the 2.5 and 97.5 percentiles).....126

Figure 6-1. Illustration of intra-operative application of the Signature PSI (Biomet Inc.) to the knee joint.....134

Figure 6-2. Flow of patients through the Signature UKA study. 138

Figure 6-3. Screenshot of the Signature PSI planning software.....139

Figure 6-4. The Signature patient-specific guides and accompanying bone models. Left image is the tibia and right image is the femur.....139

Figure 6-5. Intra-operative use of Signature PSI. Top two images demonstrate their use on the femoral side. Bottom two images demonstrate use on the tibial side.....140

Figure 6-6. Radiographic assessment of component alignment (A = Femoral component varus/valgus angle; B = Tibial component varus/valgus angle; C = Femoral component flexion/extension angle; D = Tibial component poster/inferior tilt angle).....143

Figure 6-7. Bar chart demonstrating the differing bearing sizes implanted in each group.....147

Figure 6-8. Thickness of resected bone expressed as the difference between the "planned" and "actual" resection (the box represents the IQR, the bold line in the box represents the median, the whiskers represent the 2.5 and 97.5 percentiles, and circles are outliers). 149

TABLE OF TABLES

Table 2-1 Delphi survey results: Cronbach's α values and descriptive statistics for "Impact of surgical steps of Oxford UKA on outcome" (n=34). (Likert scale: *1=Little impact; 2=Moderate impact; 3=Substantial impact*).....43

Table 2-2 Summary of the overall OATSS ratings that the two assessors awarded to each of the OUKA procedures (maximum possible OATSS = 90). 47

Table 3-1. Delphi survey results (Cronbach's α values and descriptive statistics) for identifying important parameters for analysis during post-operative radiological assessment of OUKA (n=32). (Likert scale: *1=Not important; 2=Occasionally important; 3=Important; 4=Of vital importance*).....63

Table 3-2. Definitions of important post-operative radiographic parameters and how they are assessed. 65

Table 3-3 Details of the patient sub-groups in each cohort (note that comparisons were made between the radiographic parameters of each sub-group *within* a cohort, i.e. no comparisons were made *between* cohorts). 72

Table 3-4. Influence of sub-optimal radiographic parameter values in column 'A' on risk of revision relative to values in column 'B' for Cohort 1 (e.g. a value of 2 implies that the risk of failure to those in category A is twice the risk of failure to those in category B). Instances where there are no cases in either column A or B are labelled as 'Inestimable'; ** indicates RR values that were statistically significant. 75

Table 3-5. Influence of sub-optimal radiographic parameter values in column 'A' on risk of revision relative to values in column 'B' for Cohort 2 (e.g. a value of 2 implies that the risk of failure to those in category A is twice the risk of failure to those in category B). Instances where there are no cases in either column A or B are labelled as 'Inestimable'; ** indicates RR values that were statistically significant. 77

Table 3-6. Influence of sub-optimal radiographic parameter values in column 'A' on risk of revision relative to values in column 'B' for Cohort 2 (e.g. a value of 2 implies that the risk of failure to those in category A is twice the risk of failure to those in category B). Instances where there are no cases in either column A or B are labelled as 'Inestimable'; ** indicates RR values that were statistically significant. 78

Table 4-1: Responses of the survey participants with regards to how difficult it is to teach the key steps of the Oxford unicompartmental arthroplasty procedure (n=33). Percentages are given in parentheses.....99

Table 4-2: Demographics of participants evaluating the usability of the ViTAL cognitive simulator. 101

Table 4-3: Opinions of participants with regards to the usability of the ViTAL cognitive simulator. Rating based on 5-point Likert scale (1 = strongly disagree to 5 = strongly disagree). 102

Table 4-4: Demographics of the medical students taking part in the randomised controlled trail to assess the effectiveness of the ViTAL cognitive simulator. Results are not statistically significant. 103

Table 5-1: Comparison of demographics between orthopaedic trainees in the Standard Training and ViTAL Training cohorts. All values were not significantly different between groups ($p > 0.05$). 123

Table 6-1. Summary of the study subjects' demographics. 146

Table 6-2. Comparison between the component sizes implanted and those planned. 148

Table 6-3. Proportion of radiographic parameter outliers in the Signature Instrumentation and Conventional Instrumentation groups (comments in brackets indicate the direction of the outliers). 150

Table 6-4. Summary of the radiographic parameters in the Signature Instrumentation and Conventional Instrumentation groups (** denotes statistically significant difference; SD = Standard Deviation). 151

Table 6-5. Summary of the secondary outcome parameters. (** denotes statistically significant difference) 152

Table 6-6: Summary of the surgeons' overall impression of the PSI system. 153

1. GENERAL INTRODUCTION AND LITERATURE REVIEW

1.1. INTRODUCTION

"Surgery" is derived from the Greek word *kheirourgos,* meaning "working by hand". In its crudest form, it can be traced as far back as the beginning of the human race. Although the primitive methods of the past have long been replaced, an emphasis on learning and teaching the surgical craft has clearly been alluded to throughout history. One of the most relevant examples is evident within the Seal of Cause (or Charter of Privileges) document that marked the formal incorporation of the Barber Surgeons of Edinburgh as a Craft Guild of the city in 1505[1]. In outlining the privileges and duties of the Barber Surgeons, the document states: *"... that no manner of person occupy or practice any points of our said craft of surgery... unless he be worthy and expert in all points belonging to the said craft, diligently and expertly examined and admitted by the Maisters of the said craft and that he know Anatomy and the nature and complexion of every member of the human body... for every man ocht to know the nature and substance of everything that he works or else he is negligent".* Thus, it appears that even the comparatively crude surgical practices of the 16th century Barber Surgeons placed an emphasis on the need to acquire the necessary knowledge and skills in order to become a competent and safe practitioner.

1.2. SURGICAL COMPETENCE

The notion of "Surgical Competence" has been the focus of much debate over recent years[2-4]. However, there is still no universally accepted definition of this term, or the exact criteria that determine whether or not someone has achieved it. In the context of surgery, it has been described as "… the ability to successfully apply professional knowledge, skills, and attitudes to new situations as well as to familiar tasks"[5]. Clarification of what constitutes surgical competency is necessary because it will provide the framework for designing structured surgical training programmes and curricula. Furthermore, it will facilitate the development of assessment methods that can determine whether or not surgical trainees meet the necessary competencies. Efforts have been made by a number of educational bodies to define the core competencies that are expected of doctors. One of the most widely accepted is the CanMEDs model that was proposed by the Royal College of Physicians and Surgeons of Canada[6]. The seven key roles expected of a competent doctor were: Medical Expert, Communicator, Collaborator, Manager, Health Advocate, Scholar, and Professional. Another commonly used model is the six competencies proposed by the United States Accreditation Council for Graduate Medical Education (ACGME)[7]. These are Patient Care, Medical Knowledge, Practice-based Learning and Improvement, Interpersonal and Communication Skills, Professionalism, and System-based Practice. While the underlying principles and frameworks for both the CanMEDs and ACGME competencies are similar, the specific characteristics relating to surgery can be found under different headings. For example, CanMEDs refers to technical skill under the "medical expert" competency, whereas ACGME discusses it under "patient care". These models therefore provide a useful conceptual framework for

assessing the skills, knowledge and attitudes required of a competent surgeon. However, overlap exists between these competencies, and one must remember that surgical competence remains a multidimensional and complex concept. A more pragmatic method may be to classify the key qualities of a competent surgeon under the broad headings of "technical" and "non-technical" (personal and cognitive) attributes. This is supported by the results of a Delphi survey carried out by Sir Alfred Cuschieri and colleagues[8] to determine the opinion of master surgeons (Delphi methodology is later described in Section 2.1.2). Accomplishment of these attributes is therefore essential for the development of surgical expertise.

1.3. SURGICAL EXPERTISE

Surgical expertise and the processes involved in its development have always attracted considerable attention[9-13]. Understanding and defining the often complex and interdependent abilities of surgical experts, and the means by which these attributes are achieved, remains one of the major objectives for those interested in surgical education and training. Surgical experts are not mere technicians who are capable of performing challenging surgical procedures. In addition to their superior technical abilities, they possess a number of cognitive (knowledge depth and judgement), and interpersonal abilities (communication skills, teamworking, professional values) that by far surpass those of the "average" surgeon. This concept is best elucidated by Ericsson, who proposed that most professionals reach a stable, average level of performance and maintain this status quo for the rest of their careers, whereas experts are able to go further: "The fundamental theoretical

challenge is to explain how most people and professionals reach a stable performance asymptote within a limited time period, whereas expert performers are able to keep improving their performance for years or decades"[14]. This is supported by the earlier work of Hatano and Inagaki who put forward the concept of two types of expertise, namely "routine" and "adaptive"[15,16]. Although both types of expert have mastered complex and sophisticated domain-specific skills that they can apply effectively, routine experts are not as well equipped to deal with novel problems. In such situations they persevere with their existing routine, i.e. what they know best, in an attempt to make the problem fit their skills. Adaptive experts, on the other hand, have the advantage of possessing greater flexibility and creativity at problem-solving[17]. Thus, when learning, routine experts focus on improving efficiency by refining specific aspects of established routines, whereas adaptive experts will use a new problem as an opportunity to employ innovative ways of pushing themselves further. This may at times affect the learning curve of adaptive experts and these individuals may take longer to master some complex tasks compared with routine experts. However, the adaptive experts will be better equipped to deal more effectively with complex and unplanned situations in the future. Surgical expertise is therefore better defined as a form of adaptive expertise.

1.3.1. Automaticity

Cognitive psychology research has revealed that experts (including surgical experts) accrue a vast "arsenal" of domain-specific memory skills, which can be stored in their long-term memory and be accessed effortlessly during problem solving and task execution[18,19]. Such attributes enable surgical experts to anticipate potential adverse events and perform complex tasks with so-called

"automaticity"[20-22]. This is evidenced by their ability to perform primary tasks with an apparent absence of intentional effort, with the associated capability to multitask with relative ease. Evidence of automaticity has been demonstrated in simulation-based laparoscopy tasks, during which experts were better at performing a laparoscopic task and a secondary visual-spatial task compared to non-experts[23]. Furthermore, the same researchers were able to demonstrate the benefits of simulation-based training in a subsequent study by the fact that the non-experts, through repeated practice, could improve at the simulator and secondary visual-spatial tasks[24].

1.3.2. Error detection ability and diagnostic ability

Continuing along this theme, studies during open surgical procedures have shown that the ability to detect surgical errors is strongly correlated with increasing surgical expertise[25]. Experts are, therefore, both effective and efficient at using forward reasoning to reach diagnoses, decide treatment strategies, solve operative dilemmas, and make fewer cognitive errors. Hsu et al.[26] demonstrated in a simulation study that highly experienced surgeons were able to reach "automaticity" in surgical tasks, evidenced by their ability to "do more than one thing at a time", by simultaneously undertaking an arithmetic task and a simulated laparoscopy procedure. The knowledge of these experts has been described as being "encapsulated", with memory organised around practical schemas, resulting in an improvement in diagnostic accuracy and quality as expertise develops[27]. It is clear that such reasoning strategies form a core component in many of the non-technical skills required of surgeons, such as diagnostic and decision-making skills. Schmidt and Rikers[27] argue that acquiring expertise requires the formation of "illness scripts", which are a product of extensive experience with patients

superimposed on a formal knowledge structure. Acquired through clinical experience, such skills explain the apparent observation that individuals who have "vast experience under their belt" often excel in surgical procedures.

1.3.3. The influence of surgical volume

Traditionally, expertise in the field of medicine has been closely linked to the relative experience of a doctor[14], and numerous studies have investigated the relationship between volume and clinical outcome[28]. Large studies including systematic reviews, across a wide range of surgical specialties, suggest that there is a clear and consistent relationship between both surgeon and hospital volume with outcome[28-32]. The published studies suggest that high volume surgeons (ideally within a high volume centre) have superior patient outcomes, including decreased length of stay, complication rates and mortality. There is further support for this finding in the orthopaedic literature[33,34], in particular for total knee arthroplasty[35,36] and UKA (unicompartmental knee arthroplasty)[37]. These associations may well be explained by the consistent performance that results from achieving expertise in a domain such as surgery. However, it is apparent that volume alone does not account for skill level among surgeons, as variations in performance have been shown even among high volume surgeons. Thus, sheer exposure to clinical cases cannot explain why some individuals excel to become surgical experts whilst others remain stagnant.

1.3.4. Deliberate practice

The concept of "deliberate practice" was described by Ericsson in his seminal work relating to the development of expert skill levels in domains such as sports, music, chess, and medicine[14]. This involves the repetitive performance of a

cognitive or psychomotor skill in a specific domain, coupled with informative feedback on performance. Ericsson proposed four key conditions that must be met during this process[38]. Firstly, the individual concerned needs to be *motivated to learn* and improve. Secondly, he or she must be provided with a *defined goal* which takes into consideration any pre-existing knowledge or skill. Thirdly, the individual must be provided with immediate and constructive *feedback* on performance. Finally, the individual needs to have the opportunity to *practice repeatedly over an extended period* of time. Thus, although many quote Ericsson's work, which suggests that 10 years or 10,000 hours of "sustained deliberate practice" are necessary to achieve expertise in a given domain[39], it is important to bear in mind that (as in many other walks of life) it is not just the *quantity* but also the *quality* of surgical training that influences the development of a surgeon.

1.4. What kind of practice makes perfect?

1.4.1. Distributed practice and the spacing effect

Educational theory suggests that the learning of a new skill can be strongly influenced by the type of practice schedule that is employed to learn that skill[40]. When learning a new task (whether cognitive or psychomotor), practice schedules can either be "massed" together in continuous blocks with little or no rest, or "distributed" apart with relatively long periods of rest between them. This theory has been studied extensively, and the sheer size of the effect and its reliability is impressive. There is now an abundance of evidence suggesting that tasks can be learned more effectively when learning sessions are distributed over a period of time[41]. The primary reason for this appears to be due to *reactive impedance* – where factors such as fatigue and boredom have a negative impact on learning.

Surgical educators have, therefore, looked to capitalise on this concept in order to aid memory and skill retention during complex surgical tasks[42]. The benefits of the "distributed practice" method of training have subsequently been demonstrated in several randomised controlled studies in other surgical subspecialties[43-45]. It has been postulated that this enhanced performance is due to the consolidation of learned skills[46], as well as the fact that distributed practice enables the individual to cognitively prepare for future practice sessions[47].

1.4.1.1. Contextual interference

First described by Battig when investigating verbal learning[48], *contextual interference* (CI) is one of the most frequently studied learning phenomena. Battig defined the CI effect as "the effect on learning of the degree of functional interference found in a practice situation when several tasks must be learned and practiced together". His work suggested that, compared to blocked practice, the high levels of CI encountered when a task is practiced in a random order resulted in a lower rate of skill acquisition, but that this random practice enhanced skill retention and transfer due to enhanced information processing. This concept has since been shown to be relevant to other forms of learning, including psychomotor tasks[49,50]. However, studies suggests that the beneficial effects of *random practice* are most relevant to simple skills, and that complex skills are best learned in *blocked practice*[51]. Further studies have been able to extrapolate these findings to the field of surgery[42,52], with the conclusion that surgical procedures that are composed of several discrete components should be practiced as a whole, but if training circumstances (such as time and resources) necessitate part practice of skills, then this should be organised in random order to maximise the learning opportunity.

1.5. LEARNING IN THE DOMAIN OF SURGERY

1.5.1. Rasmusson's model of human behaviour during complex tasks

The manner in which complex tasks are organised and processed by individuals has been studied extensively in cognitive psychology, with Rasmussen describing three domains based on the mental processing behaviours employed at each level[53]. He classified human performance during complex tasks into "skills-based, rule-based and knowledge-based" behaviours. This model has been successfully utilised by other surgical subspecialties as a framework for studying the complexities of performing surgical procedures[54,55]. Skill-based behaviours require almost no conscious effort, comprising very familiar actions performed in comfortable surroundings with the surgeon being on "autopilot". An example of this in surgery is the suturing of a simple wound. Rule-based behaviours are familiar to the surgeon, follow certain procedural steps and have a logical progression leading towards a known end goal. Many tasks in orthopaedic surgery (in particular primary joint arthroplasty) are composed of a specific number of procedural steps. An example during OUKA surgery is the principles for milling the medial femoral condyle and equalisation of the 20º and 90º flexion gaps, which involves flexion gap measurements followed by sequential milling based on a simple formula. Knowledge-based behaviours tackle situations that are new, unfamiliar and more complex, thereby demanding a significantly higher level of cognitive input. They depend on the performer's fundamental knowledge, diagnostic, and analytical skills, with examples related to UKA surgery being the intra-operative management of a tibial plateau fracture, or post-operative management of unexplained pain. It is important to note that some scenarios may shift from level to level depending on the individual. For example, in certain

situations, a rule-based task for one individual may be more knowledge-based for another depending on their level of experience and surgical expertise in that procedure. Rasmussen's model can be applied to surgery in order to enhance training and assessment methods. By determining the kind and amount of skill-, rule-, and knowledge-based behaviours required for a particular surgical procedure, it is possible to design more effective teaching and assessment methods for it.

1.5.1.1. Learning psychomotor skills

Psychomotor skills involve the complex, precise, and coordinated motor movements that require cognitive input. They are considered to be one of the key attributes of a competent surgeon and their development is paramount in surgical training. Fitts and Posner's 3-stage theory of psychomotor skill acquisition has been well studied[56] and is of particular relevance to surgery. These are the *cognitive*, *associative*, and *autonomous* stages. During the cognitive stage, the learner intellectualises the task and must become aware of what the task entails. The learner understands how to execute the various components of the task, however, this alone will not be sufficient to enable the learner to perform the task competently. The next stage is the associative (integrative) phase, during which the task is practiced repeatedly, whilst receiving feedback. This feedback process, also known as *knowledge of results*, is key to the learning process[57]. The learner is, therefore, able to improve his/her performance through visual, tactile or auditory feedback that helps critique any errors. The final stage is the autonomous stage, during which the task can be smoothly performed with very little or no cognitive input. This is reflected by the automaticity with which surgical experts perform

routine surgical procedures. One of the first surgeons to apply this theory to teaching and evaluating surgical skills was an orthopaedic surgeon named Joseph Kopta[58,59]. His elegant and systematic analysis of this topic helped provide one of the first skill assessment scoring systems in surgery.

Another model that has relevance to surgery is the *cognitive apprenticeship* model[60,61]. In this model, the mentor decontextualises the skill that is being taught to the apprentice, so that it can subsequently be applied to solving different problems. The mentor provides the apprentice with *modelling*, *coaching*, and *scaffolding*. Modelling is used to provide the apprentice with explicit demonstration and explanation of the skill, coaching involves the observation and provision of feedback on performance, while scaffolding involves the provision of support where necessary. The apprentice must in turn *articulate, reflect*, and *explore* in order to improve. Articulation requires the apprentice to articulate the knowledge and cognitive processes necessary for the task. During reflection, the apprentice compares their own problem-solving ability to that of their mentor or other students. The final stage of exploration involves the apprentice researching and testing new hypotheses on their own in order to solve new problems. The cognitive apprenticeship model was primarily developed to solve the traditional *apprenticeship* model, which only consisted of *observation*, *coaching*, and *practice*. This model was concerned with the learning of skills in one specific context and involved the gradual development of skills until they could be performed independently.

1.5.1.2. The surgical learning curve and innate ability

The acquisition of a new surgical skill is characterised by a ‘‘learning curve’’ (a phenomenon in which skill at performing a psychomotor task improves with repetition)[62-64]. Typically, this improvement is initially rapid and slows down as the curve reaches a plateau. The initial rate of learning of a surgical task is a measure of how quickly a surgeon reaches a certain level of performance, whereas the plateau, or asymptote, is a measure of the level at which the performance stabilises. Surgical procedures that are more challenging will have steeper learning curves. Previous studies have demonstrated that surgeons with greater surgical experience can achieve more rapid stabilisation of their learning curve[65]. The effect of training on the ability of a surgeon to stabilise his or her learning curves has been clearly demonstrated in orthopaedic surgery[66-69], as well as in other surgical subspecialties[70-72].

The innate ability of a surgeon has a strong influence on the learning curve. This was demonstrated by a recent orthopaedic study showing variations in the arthroscopic ability of medical students who had never been exposed to any type of endoscopic surgery[66]. The findings demonstrated that the novice subjects clustered into three groups of surgical ability; a group who were innately gifted with arthroscopic skills almost from the outset of the study, a group who were able to reach competency with repeated practice on a simulator, and a group who could not achieve basic competency despite repeated practice. Furthermore, in a subsequent follow-up study, the same group was able to show that there are some novices who could not even improve their arthroscopic performance with focused training[67]. These findings are supported by another study by Grantcharov and Funch-Jensen, who demonstrated a similar finding in surgical residents[73]. This

suggests that some individuals do possess innate abilities that set them apart from the rest.

1.6. SURGICAL EDUCATION: PAST AND PRESENT

1.6.1. Surgical apprenticeship and the Halstedian era

Surgery has traditionally been taught through an apprenticeship model. Prior to the formation of the 'schools of surgery' (which have subsequently become institutions such as The Royal College of Surgeons of England), skills were passed down from friends or acquaintances with little structure in training[74]. Even throughout the 20th Century, there was no real systematic approach to teaching surgical skills. One of the key changes to 21st Century surgical education occurred in 1904, when the German-style residency training system was introduced to the Johns Hopkins Hospital in Baltimore by Sir William Halsted[75]. With a strong emphasis on graded responsibility and exposure to a sheer volume of cases, this model remained the cornerstone of surgical training for decades and has produced many surgical masters. The surgeons who trained under this system accrued substantial surgical experience but the majority of their work, particularly out of hours emergency surgery, was unsupervised. A concern was that patients were sometimes harmed due to inexperienced surgeons operating beyond their abilities without adequate supervision[76].

1.6.2. Recent challenges facing the delivery of surgical education

In the past two decades there have been substantial changes to the infrastructure and delivery of surgical education – especially in Europe and the United States. One major change has been enforced reductions on surgical trainees' working hours[77-79]. In the UK, estimates suggest that the average time available to train a

surgeon has been reduced by 80%[80]. This has resulted in a decrease in the average number of operative cases per trainee, and fewer opportunities for learning surgical skills in the operating room[81,82].

Another factor that negatively impacts on modern-day surgical training is pressure from service-driven healthcare systems. The current target-driven culture of the UK National Healthcare System (NHS) is at odds with surgical training, given the additional resources and time required. It is generally accepted that trainees take longer to perform cases in comparison to consultant surgeons, and the financial implications of providing this educational opportunity have been documented both in the UK[83] and abroad[84]. In addition, there has been the growing trend of Independent Sector Treatment Centres (ISTCs) "cherry-picking" low-risk and straightforward surgical cases and leaving NHS hospitals with more complex ones. This has had a detrimental effect on surgical training by reducing the educational opportunities available to surgical trainees[85,86].

These problems were exacerbated by the reforms that were introduced in 2004 under the Modernising Medical Careers (MMC) programme[87]. There was an effort to improve and streamline training by moving from an apprenticeship-based model to one that is competency-based. This attempt to train competent surgeons in a shorter time period necessitates further emphasis on providing focused and high quality training. As previously mentioned, Ericsson's theory of 'deliberate practice' is key to the development of surgical expertise, and adequate training time is important to facilitate this[39]. This can only be achieved through innovative and efficient training methods - such as simulation-based training. This must be coupled with robust, yet formative, assessment methods to optimise learning and ensure that future surgeons have the necessary competencies to fulfill their

duties[88-90].

1.7. ASSESSMENT OF SURGICAL SKILLS

1.7.1. What are the basic requirements of assessment tools?

Over the past two decades, considerable efforts have been made to develop better methods of assessing surgical skills[88,91]. It is important to develop assessment methods that are feasible and can also be generalised, otherwise they will never be incorporated into surgical curricula. In addition, the metrics used must be both valid and reliable.

1.7.2. Validity

Validity is a fundamental property of a test or assessment tool and is concerned with whether or not it measures what it purports to measure[92]. Validity should not be thought of as a binary concept, but as a spectrum. Rather than being merely present or absent, there are degrees of validity, determined by the weight of supporting evidence available for that test. Proving perfect validity for any test is probably unachievable in the real world. Validation studies aim to provide sufficient evidence to support the construct as providing a true measure of what is tested within a specific context. Various components of validity (with reference to an assessment tool) have been described and are summarised below:

Construct validity is the overarching concept supported by all other forms of validity. It is the degree to which an assessment tool measures the attributes of performance it is designed to measure, such that it is able to differentiate between novices and experts.

Content validity is established by showing that the components of a test are relevant and represent the skills the evaluator is interested in. It ensures that the steps within the task are thought out and linked. This form of validity is relatively subjective and often relies on expert opinion.

Face validity is a subjective measure of how closely an assessment tool or task resembles the situation in real life and is usually measured through questioning experts. This is often a basic prerequisite of designing an assessment tool or task. A high degree of face validity can positively influence the acceptance by end-users such as trainee surgeons.

Concurrent validity is the extent to which the results of an assessment tool correlate with the "gold" standard tests known to measure the same skill domain. High correlation between different assessment tools indicates good concurrent validity.

Predictive validity is the extent to which future performance can be predicted by the test score.

1.7.3. Reliability

Reliability refers to the consistency of an assessment method[93]. It is the measure of the reproducibility of test scores obtained from an assessment given multiple times under the same conditions. All assessment involves taking a sample of an individual's knowledge or performance and making inferences about that data to reach a conclusion about the individual's true ability, of which a reliable test must give a fair reflection. The greater the difference between the test result and the individual's true ability, the less reliable the assessment. In the context of surgical skill assessment, this is usually evaluated using inter-rater reliability and internal

consistency. *Inter-rater reliability* ensures that there is good agreement between assessors of an individual's performance. Two forms exist: *interobserver* reliability measures the agreement between different assessors for a given test, whereas *intraobserver* reliability determines the variability of a single assessor's marks for the same test on different occasions. Because most skill assessment tools rate performance using more than one single domain, a good measure of performance shows good *internal consistency* among the different test items used for assessment. The assumption here is that a skilled surgeon will rate highly (although not identically) throughout all the different domains.

1.7.4. Current methods of assessing surgical skills

1.7.4.1. Observational tools for skill assessment

The modern-day methods of evaluating and quantifying surgical skill acquisition by means of observational assessment began in 1997, with the introduction of the "Objective Structured Assessment of Technical Skill" (OSATS) by Reznick and colleagues in Toronto[94]. OSATS consists of assessors using two methods to assess the surgical skill of a candidate. The first is a task-specific checklist that assesses a specific procedure by breaking it down into smaller surgical steps. The second is a "Global Rating Scale" (GRS) that provides a qualitative outcome measure of skill domains necessary for performing a procedure (e.g. instrument handling, flow of operation, and quality of final product). A GRS is not necessarily linked to a specific task and therefore is more generic than a task-specific checklist. GRSs have been found to have better construct validity than task-specific checklists[95-97]. OSATS can be performed at the time of the procedure or at a remote time and location with video recordings - a feature that makes it an economical and flexible

assessment tool. They have high reliability and validity, but their disadvantage is that they require trained assessors and significant time; therefore not every institution may readily take them up.

Although initially designed and validated for open general surgical procedures, research in this area has grown significantly over recent years and many studies have since use adaptations of GRS, task-specific checklists or a combination of the two, for observational assessment of surgical skill in many surgical subspecialties[91,98]. Their success has resulted in their adoption by educational bodies such as the UK's Intercollegiate Surgical Curriculum Project (ISCP)[99], who developed work-based assessment tools such as Procedure-Based Assessments (PBAs) and surgical Direct Observation of Procedural Skill (DOPS)[100].

Within orthopaedics, several observational assessment tools have recently been developed and validated to evaluate technical performance in arthroscopic surgery[101-107]. However, no such tools have yet been validated for arthroplasty procedures.

1.7.4.2. Motion analysis and measurement of force patterns

Motion analysis is a 3-dimensional (3D) dexterity analysis system that measures the efficiency of hand movements. It has been extensively validated as an objective tool for assessing technical skill in orthopaedic surgery[67,69,108] and other surgical subspecialties[109,110]. The electromagnetic signal that is emitted from the transmitter is received by sensors on the back of a subject's hands (Figure 1-1). These motion tracking devices measure the total time taken to complete the procedure, total path length travelled by the subject's hands, and total number of hand movements during the procedure. The three metrics are inversely

proportional to surgical ability, thus an individual with greater technical skill would be able to execute a given surgical task in less time and with a shorter total path length and fewer hand movements.

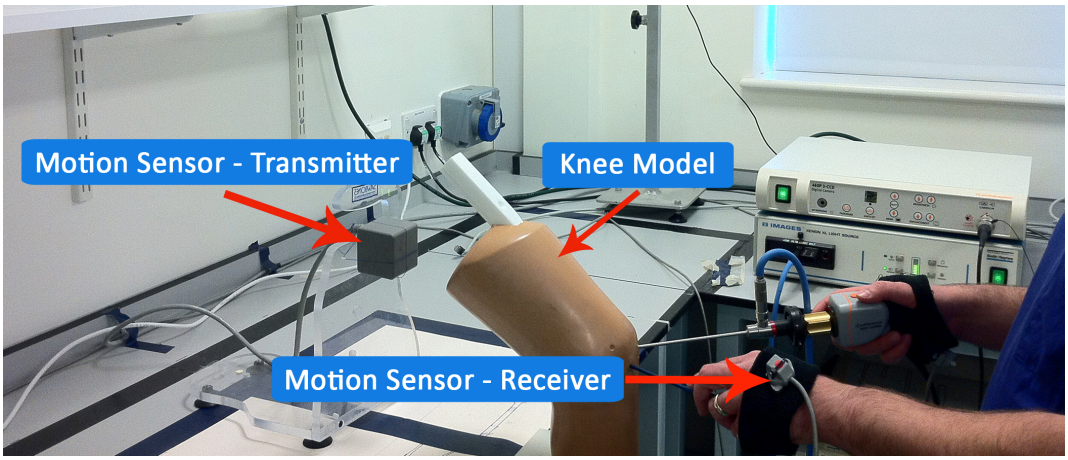

Figure 1-1 Motion analysis system being used during a simulated arthroscopy task on a phantom knee model.

Although less frequently used than motion analysis, the evaluation of force patterns has been used as a determinant of arthroscopic skill. The rational for this lies in the fact that iatrogenic damage can occur if excessive force is used when manipulating arthroscopic instruments within a joint cavity. Several studies have consequently demonstrated correlations between force patterns and the level of arthroscopic experience[111-113]. As with motion analysis, this type of assessment can only be carried out in the surgical skills laboratory setting.

Both these assessment modalities have imitations. Firstly, in their current forms, they can only be used in the surgical skills laboratory setting and are not transferable to the real operating theatre. Furthermore, they do not assess other important domains of surgical expertise, such as decision-making, knowledge, and communication skills.

1.8. ERRORS AND ADVERSE EVENTS IN SURGERY

The dynamic, pressured and high-risk environment that often surrounds surgery is a source of surgical errors, which can negatively impact on patient care. Numerous studies have demonstrated that the very system and healthcare professionals who are responsible for providing care to patients can actually be responsible for causing them harm[114-119]. The cause can vary from a single error by an individual surgeon, to the breakdown of several safety systems within the healthcare institution[120,121]. More importantly, a recent systematic review into the nature and incidence of in-hospital adverse events suggested that almost half of all adverse incidents were preventable[122]. Although it is human nature to make errors, the ultimate goal when providing patient care is to minimise or even abolish such events. The seminal report by the Institute of Medicine in 1999 called *To Err is Human: Building a Safer Health System*[123] highlighted the need for significant system-wide changes in order to improve patient safety in medicine. Consequently, over the past two decades, the healthcare profession has turned to so-called "high reliability organisations" as a model for improving its safety management systems[124,125]. These are organisations that are inherently high-risk but have very low failure rates due to their safety management systems. They include the military, commercial aviation and the nuclear power production industry. As a result, many aspects of the training systems, organisational principles, and policies that are used in these industries have been employed to improve training, performance and safety within healthcare.[126-128]

The focus on surgical safety and competency further intensified following a number of well-publicised incidents both in the UK[129,130] and abroad[131,132]. The public enquiries that followed incidents such as the "Bristol Heart Scandal"

highlighted poor surgical performance by the surgeons concerned[129]. The resultant increase in public and peer scrutiny has meant that the need to demonstrate high standards of surgical care and performance has never been so great. Many people in modern-day society will no longer accept that surgical training could result in patient morbidity or mortality. The institutions responsible for delivering surgical training are consequently under pressure to demonstrate that they are providing comprehensive, structured, and evidence-based programmes that pose no risk to patients and produce surgeons of the highest calibre.

1.9. RECENT METHODS EMPLOYED FOR IMPROVING SURGICAL EDUCATION

1.9.1. Computer-aided learning and web-based educational resources

In both medicine and surgery, the acquisition of 'knowledge' is one of the central educational goals - with the other central objectives being the acquisition of 'skills' and 'attitudes'[133]. Advances in technology mean that the traditional methods of conveying knowledge with books and lectures are now being supplemented, or at times even superseded, by readily available internet-based learning resources. Self-directed learning has been highlighted as an important aspect of adult learning theory, which empowers individuals to assume responsibility for their own learning[134]. E-learning tools can be important in aiding this process, and the application of sound educational principles during this process can facilitate deep learning.

1.9.2. Simulation-based surgical education

Simulation has been described as "a technique- not a technology- to replace or amplify *real experiences* with *guided experiences* that evoke or replicate

substantial aspects of the real world in a fully interactive manner"[135]. The most important aspect is the manner in which this technique is applied, rather than the actual complexity of the technology involved. Simulation-based surgical education is a relatively new paradigm that has attracted considerable attention over the past decade. It provides a training and assessment platform for technical and non-technical skills outside the real operating theatre. Use of technology to provide simulation-based education can be employed across wide areas of the surgical curriculum. This ranges from teaching basic technical skills such as venous cannulation for medical students through to complex surgical procedures carried out by surgical teams within simulated operating theatres. Regardless of the complexity of the technology used, simulation enables learners to take centre stage in order to learn, practice and improve their skills. Surgical training must be designed around the educational needs of the learner but this is not always possible in the *real* medical world, as the needs of patients come first. Simulation-based education enables trainees to take centre stage and provides a structured and controlled setting in which surgical skills can be learned. Sustained deliberate practice within this 'safe' environment will not only ensure patient safety, but also allow specific learning outcomes to be achieved. In addition, assessments can be made using objective metrics in order to provide feedback to optimise learning.

Physical model task trainers

These simulators include artificial plastic bones and anatomical joint models (Phantoms). They can now be found in surgical skills centres of many medical schools and hospitals. Examples include plastic models used to teach wound suturing, bone plating or joint arthroscopy. Despite having a low level of fidelity (realism), they provide standardised models for skills training that are easily

accessible and relatively cheap. Other notable advantages include their portability and reusability. They do, however, have several disadvantages. Firstly, they require the presence of an experienced teacher who is able to assess and critique the performance of the learner. Such access to expert tutors is limited due to manpower shortages in current skills centres. Secondly, despite technological advances in material science that have enabled the production of realistic bench-top models, the lack of clinical context can limit their educational potential. Therefore, skills such as decision making and communication skills have traditionally been difficult to teach when using bench-top models. However, some investigators have combined the use of simulated patients (i.e. trained actors) with a variety of bench-top models to teach technical as well as non-technical skills[136]. Nevertheless, physical model task trainers alone have been effectively used in orthopaedic surgery to provide training for arthroscopy[67,137-139], arthroplasty[140] and fracture surgery[52,141,142]. One study has also demonstrated the successful transfer of improved arthroscopic skills to the real operating theatre following a period of simulator training on knee phantom models[137].

Cadavers

Along with anatomical dissection, cadaveric training has been utilised in surgical education for centuries. Improvements in embalming techniques together with the wider availability of fresh cadavers mean that they now possess the highest level of fidelity compared to all the other types of simulators, and are increasing in popularity[143]. However, their routine use will always be limited due to difficulties with their procurement, ethical issues, cost, and the theoretical risk of disease transmission.

Computer-based and Virtual Reality (VR) simulators

Technological advances have resulted in an exponential growth of computer-based and Virtual Reality (VR) simulators that can recreate many surgical procedures. The level of fidelity ranges widely: low fidelity simulators aim to improve psychomotor skills by using abstract objects, whereas higher fidelity simulators such as the ARTHROMentor (Simbionix, Cleveland, Ohio, USA) are far more sophisticated and enable learners to practice entire surgical procedures such as arthroscopic meniscectomy. Some possess haptic (touch) feedback, thus creating the illusion of "real physical contact" between the arthroscopic instruments and simulated objects. However, one must remember that the key issue is not the degree of fidelity, but whether a simulator is effective at teaching or assessing a particular domain such as cognitive or psychomotor skills. VR simulators have a number of educational benefits. They have an enormous capability for data capture in the form of various performance metrics, such as the number or types of errors made or the time taken to complete a task. This can either provide real-time feedback or can be stored for subsequent analysis and discussion with an expert tutor. The progress of the trainee can therefore be followed over a designated time period or training module. They also provide standardised and reproducible clinical scenarios that can be practised repeatedly. The major drawback of VR simulators is the cost (£80,000-£100,000).

Seymour et al.[144] were one of the first groups to demonstrate the transfer validity of a VR simulator into real laparoscopic cholecystectomy operations. Such findings have since been reproduced by other groups, and there is now convincing evidence that training on VR simulators can improve the psychomotor skills of trainees and that the skills gained can be transferred to real procedures[145-147].

Although this transfer validity has not yet been demonstrated in orthopaedic surgery, there is good evidence to support the role of VR-based simulation training – especially for arthroscopic procedures[148-153]. More complex and sophisticated work has resulted in the development of simulated operating theatres in which trainees can learn technical and non-technical skills in a realistic environment, where sustained deliberate practice can be carried out without danger to real patients[154].

Incorporating novel educational methods and technology to improve surgical training is now a necessity, especially in light of the current healthcare and socio-political pressures that demand the training of competent surgeons in shorter periods. Embracing these methods will ensure that training needs are met whilst maintaining safe standards of patient care.

1.10. KNEE OSTEOARTHRITIS AND UNICOMPARTMENTAL KNEE ARTHROPLASTY

1.10.1. The burden of knee osteoarthritis

Osteoarthritis (OA) is a degenerative condition of the joints due to progressive and irreversible loss of articular cartilage[155]. The main symptoms are significant pain, stiffness and deformity. The burden of OA is huge, and it is the most common joint disease affecting just over 8 million people within the UK[156]. Apart from having a significant impact on the individual concerned, OA places an enormous economic burden on society due to the direct costs of its treatment, and also indirect costs such as loss of working days[157,158]. OA of the knee joint (Figure 1-2) is very common with epidemiological studies estimating that painful knee OA associated with disability affects up to 10% of adults aged over 55[159].

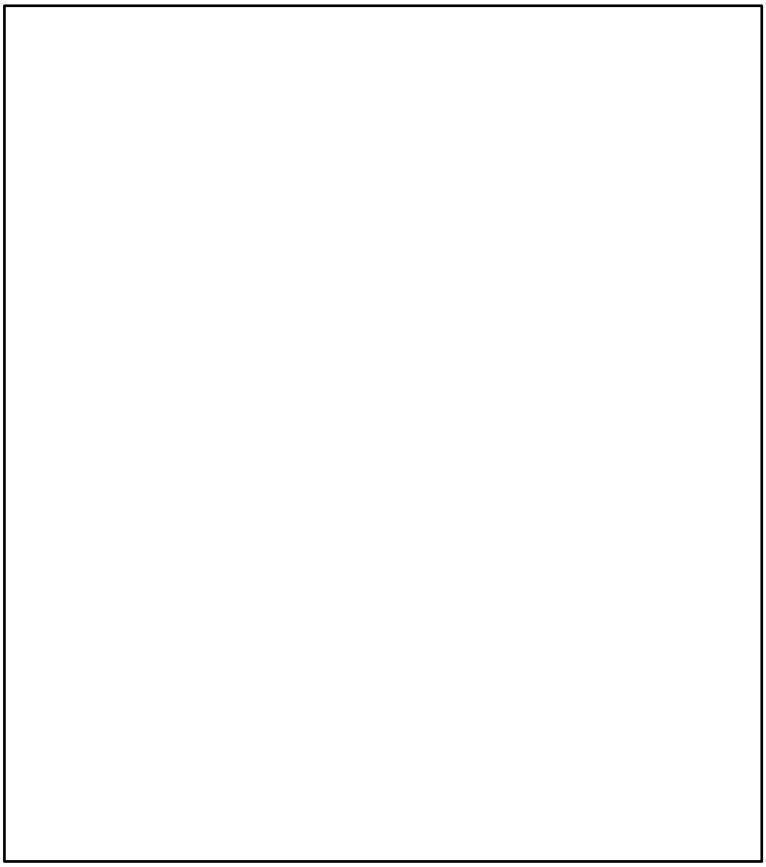

Figure 1-2 Anatomical diagram of the knee joint (Image removed due to copyright licence)

Many of these individuals will require knee replacements, and given the increasing age of the population, studies suggest that the demand is projected to grow by 673% by 2030[160]. The 2013 National Joint Registry (NJR) data show that 84,833 primary knee replacements were carried out in England and Wales during 2012, 98% of which were for OA[161]. Previous studies have demonstrated that the disease pattern is often limited to the medial compartment of the knee[162-164] (Figure 1-3).



Figure 1-3 Anteroposterior radiograph of a knee demonstrating medial compartment osteoarthritis.

Consequently, provided they meet a number of other criteria, it has been suggested that up to a third of patients may be suitable for a medial unicompartmental knee arthroplasty (UKA)[165]. The anatomical features of anteromedial osteoarthritis of the knee (Figure 1-4), described by White and colleagues[166], explain why it can be treated by UKA.



Figure 1-4 Excised articular surface of the medial tibial plateau demonstrating the classic wear pattern of anetromedial osteoarthritis.

1.10.2. Unicompartmental knee arthroplasty

Although the concept of replacing part of the knee joint was first introduced in the 1950s[167], the initial studies reporting the results of UKA were only published in the 1970s. These early results were relatively poor[168,169], but began to improve with the development of better instrumentation and refinement of the indications for the procedure. Minimally invasive UKA is now a recognised form of treatment for end-stage single compartment knee OA[170-172]. It has a number of advantages over total knee arthroplasty (TKA) that make it a more attractive option for medial compartment OA. Compared to TKA, patients undergoing UKA (Figure 1-5) have lower post-operative morbidity[173], faster recovery[174,175], lower mortality[161], preservation of normal knee kinematics[176], and increased cost efficiency[177,178].

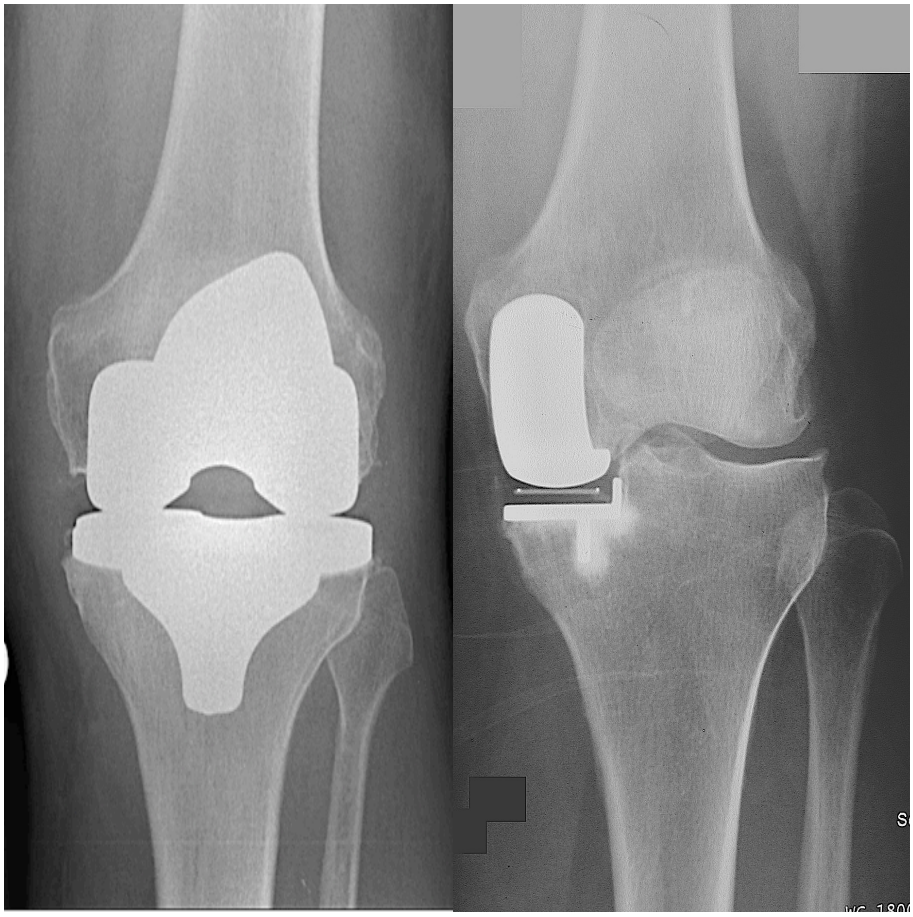

Figure 1-5 Post-operative radiographs comparing total knee arthroplasty (left) and unicompartmental knee arthroplasty (right).

1.10.3. Oxford unicompartmental knee arthroplasty (OUKA)

The Oxford UKA (Figure 1-6) is the most commonly performed UKA in England, accounting for 62% of all those performed[161]. Approximately 40,000 cases are carried out annually worldwide. Designed by Goodfellow and O'Connor, the concept of a congruous mobile bearing knee prosthesis was first introduced in 1974 as bicompartmental knee replacement[179]. Since 1982, it has been used to treat medial compartment OA of the knee[180].

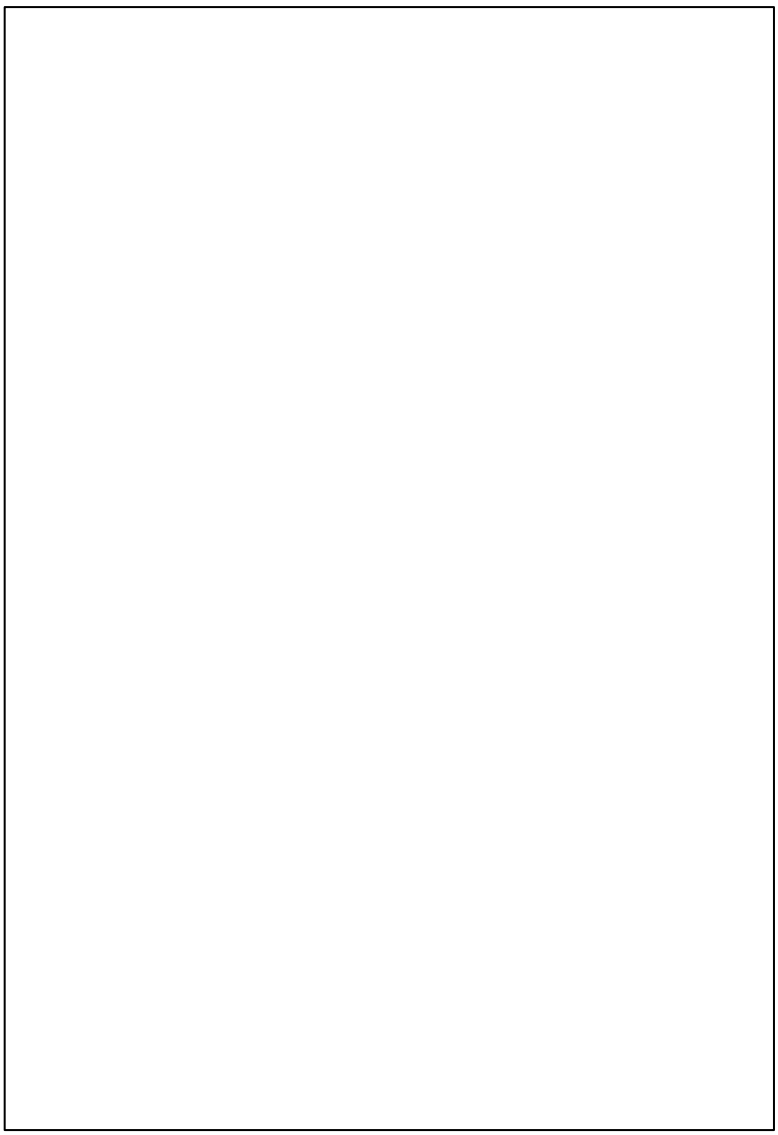

Figure 1-6 The Oxford unicompartmental knee arthroplasty. (Image removed due to copyright licence)

The earlier Phase 1 and 2 designs were implanted through a conventional open approach with dislocation of the patella. Since 1998, a minimally invasive technique (Figure 1-7) utilising specifically designed instrumentation has been introduced[165].

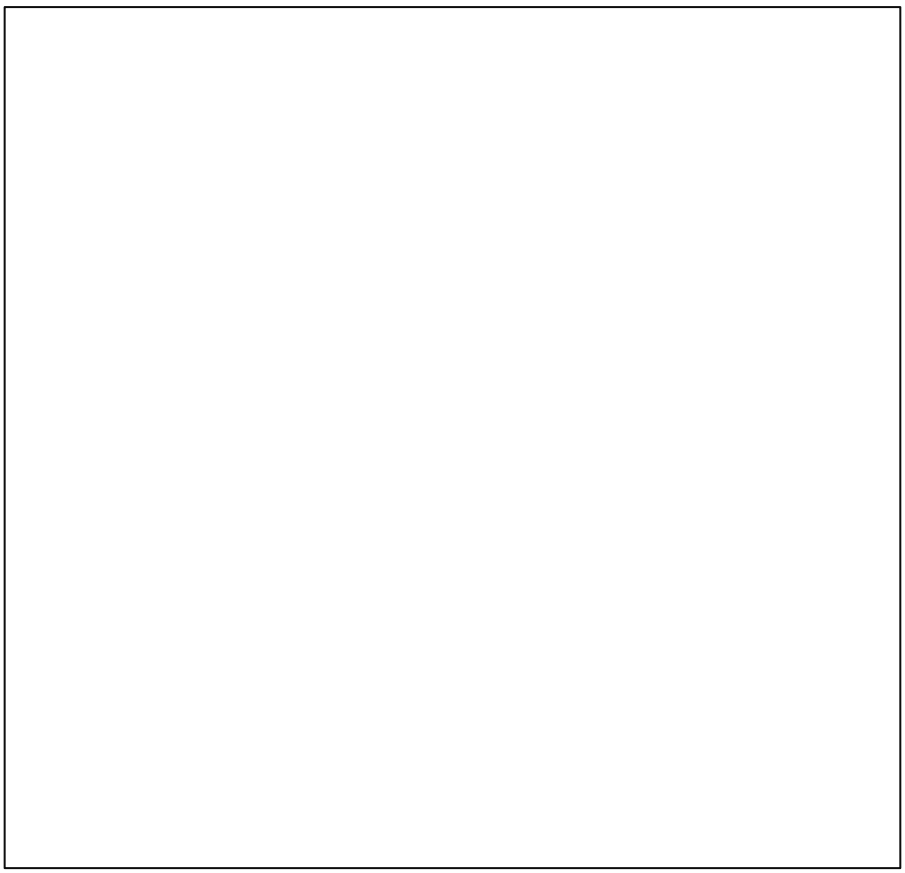

Figure 1-7 Intraoperative photograph of minimally invasive Oxford unicompartmental knee arthroplasty, demonstrating the limited view of the knee joint through the small incision that adds to the technical difficulty of the procedure. (Image removed due to copyright licence)

Numerous studies have now demonstrated excellent long-term survival rates in large cohorts of patients[181-183]. Despite the success and popularity of UKA, the NJR results indicate higher revision rates for UKAs compared with TKAs[161]. There are, however, flaws in using revision rates from joint registries as the main indicator of successful outcome after UKA[184], and joint registry studies have highlighted the important role of experience in determining successful outcomes after UKA, showing that better results are achieved for UKAs performed by high volume surgeons in high volume centres[37]. The technical difficulty of UKA and its steep learning curve is therefore thought to contribute significantly to the variations observed in surgical performance and results[185-192]. The implication of

these studies is that those aiming to undertake this procedure must be adequately trained and should perform a certain number of cases annually.

1.10.4. Current methods of teaching and improving performance during the OUKA procedure

As with many other joint arthroplasty procedures, supervised training in a real operating theatre remains the most popular method of teaching OUKA. In addition, the developers of the prosthesis have, for many years, provided surgical education through the organisation of regular instructional courses consisting of lectures, synthetic bone workshops and live demonstrations[193].

Other methods that have been used to improve surgical performance during UKA include computer-guided navigation and robotic systems. However, these technologies have a number of drawbacks including cost and time constraints, and specific complications such as tracker pin site-related problems[194-206]. These limitations, together with the lack of long-term functional outcome data, have prevented the widespread adoption of these technologies.

1.10.5. Chapter summary and rationale of this thesis

This literature review highlights the challenges that face modern-day surgical trainees and some of the methods employed by surgical educators to combat these issues. The work of other surgical subspecialties suggests that the use of novel technologies, together with the improvement of teaching and assessment methods, are key factors that can positively affect surgical training and performance. In addition, there are increasing calls for further integration of simulation-based education into surgical training. However, as highlighted by a recent review, use of these methods has been somewhat lacking for open arthroplasty surgical

procedures[207]. It is felt that many of these methods can also be utilised to improve performance during arthroplasty procedures. By virtue of its popularity, technical difficulty and variable results, the OUKA procedure is an ideal 'test case' to further explore these issues with a view to developing further solutions.

1.11. SPECIFIC AIMS

This thesis sets out to explore and answer the following questions:

i. Can assessment of surgical performance during the OUKA procedure be improved?

ii. Can novel technologies, such as VR-based simulators, improve learning of the OUKA procedure?

iii. Can novel technologies, such as patient-specific instrumentation, be used to improve surgical performance during the OUKA procedure?

1.12. THESIS STRUCTURE

Chapter 2 consists of three studies, and reports the development and validation of a procedure-specific assessment scale for evaluating surgical performance during OUKA. Study 2A uses the consensus-based Delphi methodology to determine the content of this procedure-specific assessment scale. Studies 2B and 2C are clinical studies that determine its reliability and validity in the real operating theatre.

Chapter 3 consists of two studies and reports on indirect methods for assessing performance during OUKA by means of post-operative radiological evaluation. Study 3A uses another Delphi process to determine important radiological parameters that must be evaluated post-operatively. Study 3B reports on the relationship between these parameters and the clinical outcome of OUKA.

Chapter 4 reports on the development and validation of a VR-based cognitive simulator for teaching the OUKA procedure.

Chapter 5 presents the results of a randomised controlled study conducted in a surgical skills laboratory, aimed at determining the effect of cognitive skills training (using a simulator) on the ability of orthopaedic trainees to perform the OUKA procedure on synthetic bone models.

Chapter 6 presents the results of a randomised controlled clinical study performed to investigate the effect of a novel technology (patient-specific instrumentation) on improving surgical performance during OUKA.

The concluding Chapter 7 summarises the main findings of the thesis and highlights their clinical relevance. It also proposes further ideas for future studies based on the findings discussed here.

2. DEVELOPMENT AND VALIDATION OF A TECHNICAL SKILL ASSESSMENT TOOL FOR UNICOMPARTMENTAL KNEE ARTHROPLASTY

2.1. INTRODUCTION

2.1.1. Procedure-specific assessment scales

Reliable and valid assessment tools are required to evaluate technical skill, as this is one of the most important attributes of a competent surgeon. These tools can be used to evaluate a trainee's performance during a particular procedure while providing constructive feedback where necessary. They are also important in monitoring an individual's surgical learning curve and establishing when he or she has achieved competence. As a result, several objective and evidence-based assessment tools have been developed over recent years to evaluate technical skill during open[96,208], laparoscopic[209,210], endoscopic[211], and also arthroscopic[101] procedures. These initially consisted of *Global Rating Scales (GRS)* and *Task-specific Checklists*[96]. However, one disadvantage of using GRSs is that, since they only assess global "domains" of skill (e.g. "instrument handling"), they cannot be used to give feedback on specific stages of a procedure[212]. Although task-specific checklists provide more specific information, this is only available in a binary format (e.g. "performed step satisfactorily" or "performed step unsatisfactorily"). Hence they are less sensitive and have been shown to have a low 'ceiling effect'[101]. Efforts to improve upon these methods have resulted in the development

of evaluation tools that are hybrids of GRSs and checklists, namely *procedure-specific assessment scales*[213-216].

In England, the Orthopaedic Competency Assessment Project (OCAP) initiative has developed Procedure Based Assessments (PBAs), which use task-specific checklists in conjunction with other domains (such as pre-operative planning and patient consent) for the evaluation of trainees[99,217]. However, no procedure-specific assessment scale has yet been developed and validated for a complex orthopaedic procedure. The steep learning curve of OUKA makes this procedure an ideal case for which to develop a procedure-specific assessment scale.

2.1.2. Use of a consensus-based Delphi approach

In order to assess performance during a particular procedure, it is important to define its key steps. These can then be used as components of an evaluation tool for grading surgical performance. This can be challenging, as there may be disagreement as to what constitutes each of the important steps. Furthermore, in order to achieve widespread adoption and standardisation, an assessment tool must be reflective of practice across a number of institutions. The Delphi technique, which takes its name from the Delphic oracle's skills of interpretation and foresight, was originally developed in 1948 by the RAND cooperation in USA[218]. It is a well-recognised research method that uses the opinion of experts to reach consensus on a topic about which there is uncertainty[219,220]. It is an anonymous process, through which participants express their opinions through questionnaires. The anonymity ensures that no particular member is able to exert their dominance over other group members. Responses are analysed, and through controlled feedback and repeat iterations of the modified questionnaire, group consensus is

reached. It has been used in various domains of healthcare[221,222], including orthopaedics[223,224], for consensus formation. More recently, it has been successfully used by other surgical specialities to develop simulation-based curricula[225,226] and procedure-specific assessment scales in laparoscopic surgery[227-229].

2.1.3. Reliability and validity

The essential components of a technical skill assessment tool are reliability and validity[3,88,230]. In general, the *reliability* of a technical skill assessment tool refers to the extent of its precision and reproducibility when used, either by different assessors or at different time frames. *Construct validity* is the most relevant type of validity when evaluating a technique for surgical skill assessment. This refers to its ability to distinguish between subjects with varying levels of surgical expertise.

2.1.4. Aims

The primary aim of this chapter was to develop and validate a procedure-specific assessment scale for the OUKA procedure. This was achieved using three linked studies, ultimately resulting in the development of the ***"Oxford Arthroplasty Technical Skill Score" (OATSS)***.

- **Study A**: To determine the essential technical steps of OUKA by using the Delphi methodology. This would, in turn, enable the development of a technical skill scoring system for evaluating performance during this procedure.

- **Study B**: To determine the inter-observer reliability of the OATSS evaluation tool for assessment of operative performance in real operating theatre conditions.

- **Study C**: To determine the construct validity of the OATSS evaluation tool within real operating theatre conditions.

2.2. STUDY A: CONSENSUS-BASED FRAMEWORK FOR DEVELOPMENT OF THE "OXFORD ARTHROPLASTY TECHNICAL SKILL SCORE" (OATSS)

2.2.1. Methods

2.2.1.1. Study design and participant selection

This prospective study used the classical Delphi methodology to achieve consensus amongst an international group of expert knee surgeons on the essential steps of the OUKA procedure. The participants selected were practicing orthopaedic surgeons, regarded as leaders in the field of knee arthroplasty by their peers and those within national or international orthopaedic societies (e.g. 'The Knee Society', 'British Association for Surgery of the Knee', 'Australian Knee Society'). In addition, the participants had a strong publication record in this field and were involved in training other surgeons in UKA surgery. A deliberate attempt was made to select participants from a wide-ranging geographical area, which ultimately included Europe, North America, Australia, Asia and Africa. Non-responders were sent one further email two weeks after the initial invitation. As stipulated by Delphi guidelines, each participant remained blind to the identity of the other group members.

2.2.1.2. Delphi design and determination of group consensus

The development of the Delphi survey began by identifying the steps required to perform the OUKA. These steps were outlined using orthopaedic textbooks, peer reviewed literature, and the Arthroplasty Unit faculty at the Nuffield Orthopaedic Centre (which included two surgeons involved in the design and development of the OUKA prosthesis). This information was used, in conjunction with guidelines on questionnaire design[231-234], to produce the preliminary survey. The resulting survey was piloted locally at the OUKA's design centre in order to ensure that it was user-friendly and clear. As a result, some of the survey items were rephrased or clarified. The final survey used in the first round of the Delphi process (Figure 2-1) consisted of two sections. The first section asked experts to indicate (on a 3-point Likert scale) the impact of each surgical step on OUKA outcome (1 = Little impact, 2 = Moderate impact, 3 = Substantial impact). The second section asked the experts whether or not they found the surgical step in question easy to teach their trainees. The purpose of the first section was to facilitate the development of the OATSS, whereas the second section aimed to identify areas that may require further emphasis when teaching trainees. Although the information from the second section would not be necessary for the OATSS development, the findings were needed for the later development of the VR training simulator (Chapter 4) for teaching the OUKA procedure. The results of this second section will be discussed further in Chapter 4. In addition, participants were encouraged to comment on any survey items, clarify their ratings, or add any extra steps that they felt were particularly important.

The participants who agreed to take part in this study were asked to complete the Delphi survey and return it via email to the principle investigator. The intention

was to repeat further Delphi rounds until group consensus was reached on the critical surgical steps of the OUKA procedure. In order to statistically quantify the reliability of the experts' opinions, Cronbach's α[235] was selected as a measure of internal consistency for each of the surgical steps. An α value of 0.7 (i.e. agreement between 70% of experts) was selected to indicate group consensus as this is an acceptable value of internal consistency for educational research purposes[236].

	How much impact do you feel the following steps have on the ***outcome*** of the operation? (Please tick **one** option)			Whether you find it easy ***to teach*** this step to your trainees? (Please tick **one** option)	
	Little	Moderate	Substantial	Yes	No
Patient positioning	☐	☐	☐	☐	☐
Skin incision and surgical approach	☐	☐	☐	☐	☐
Joint inspection & assessment for contra-indications	☐	☐	☐	☐	☐
Osteophyte excision	☐	☐	☐	☐	☐
Seating of tibial saw guide onto tibia and estimation of resection height	☐	☐	☐	☐	☐
Execution of the vertical cut for tibia	☐	☐	☐	☐	☐
Execution of the horizontal tibia cut and removal of tibial plateau	☐	☐	☐	☐	☐
Ensuring sufficient amount of tibial plateau is excised and further resection made if necessary	☐	☐	☐	☐	☐
Insertion of the femoral IM rod with attention to patella	☐	☐	☐	☐	☐
Assembly of the femoral drill guide & placement of the femoral drill holes	☐	☐	☐	☐	☐
Excision of the posterior femoral condyle with saw using saw guide	☐	☐	☐	☐	☐
Milling and measurement of flexion / extension gaps	☐	☐	☐	☐	☐
Estimation of the trial tibial template	☐	☐	☐	☐	☐
Preparation of the tibial plateau keel slots	☐	☐	☐	☐	☐
Selection of meniscal bearing & ligament balancing	☐	☐	☐	☐	☐
Cementation of the implant	☐	☐	☐	☐	☐
Attention to MCL / soft tissues	☐	☐	☐	☐	☐
Final inspection for bearing impingement and alignment	☐	☐	☐	☐	☐

Figure 2-1 Delphi survey to identify: i) key steps in performing Oxford Unicompartmental Knee Arthroplasty; ii) whether or not these steps are easy to teach.

2.2.1.3. Creation of the final technical skill evaluation tool for OUKA

Once expert consensus was achieved, the steps that 70% of experts regarded as having a "moderate" or "substantial" impact on OUKA outcome were included in the final technical skill evaluation tool. The resulting scoring system for the assessment of performance during OUKA consisted of 18 items deemed to be the key technical steps of the procedure. They were rated on a 5-point Likert scale. The extreme and middle points of the scale were anchored by specific descriptors to aid assessors in their evaluation of operative performance.

2.2.1.4. Statistical analysis

Descriptive statistics were in the form of means and standard deviations for all surgical steps. Cronbach's α was calculated for each surgical step in order to determine the reliability of the experts' opinions for steps of the OUKA procedure. All data were analysed with SPSS software (version 18.0; SPSS, Chicago, IL).

2.2.2. Results

A total of 56 experts were emailed and invited to participate. Thirty-four agreed to take part in the survey resulting in a 61% response rate. Nine were from the United Kingdom and 25 were from abroad.

Cronbach's α was greater than 0.7 for every surgical step within the survey, indicating that panel consensus was reached for all of these steps. Although they were given the opportunity to comment on and include additional surgical steps, none of the experts provided any further steps for inclusion in the final scoring system. Therefore, all of the surgical steps that were present in the first Delphi round were included in the final scoring system. The Cronbach's α values, Likert

scale means and standard deviations for each surgical step are summarised in Table 2-1.

Surgical Steps of Oxford UKA procedure	Cronbach's α	Degree of clinical importance (based on Likert scale)	
		Mean	Std. Deviation
Patient positioning	0.76	2.21	0.696
Skin incision and surgical approach	0.76	2.21	0.650
Joint inspection and assessment for contra-indications	0.76	2.48	0.712
Osteophyte excision	0.72	2.61	0.747
Seating of tibial saw guide onto tibia and estimation of resection height	0.75	2.73	0.626
Execution of the vertical cut for tibia	0.75	2.82	0.635
Execution of the horizontal tibia cut and removal of tibial plateau	0.74	2.73	0.626
Ensuring sufficient amount of tibial plateau is excised and further resection made if necessary	0.723	2.79	0.650
Insertion of the femoral IM rod with attention to patella	0.79	1.98	0.755
Assembly of the femoral drill guide and placement of the femoral drill holes	0.75	2.79	0.415
Excision of the posterior femoral condyle with saw using saw guide	0.75	2.58	0.614
Milling and measurement of flexion / extension gaps	0.75	2.82	0.465
Estimation of the trial tibial template	0.75	2.48	0.508
Preparation of the tibial plateau keel slots	0.73	2.52	0.566
Selection of meniscal bearing and ligament balancing	0.77	2.85	0.364
Cementation of the implant	0.74	2.70	0.467
Attention to MCL / soft tissues	0.76	2.91	0.292
Final inspection for bearing impingement and alignment	0.75	2.70	0.585

Table 2-1 Delphi survey results: Cronbach's α values and descriptive statistics for "Impact of surgical steps of Oxford UKA on outcome" (n=34). (Likert scale: *1=Little impact; 2=Moderate impact; 3=Substantial impact*)

The resulting procedure-specific assessment scale for assessing performance during OUKA was named the ***"Oxford Arthroplasty Technical Skill Score" (OATSS)*** and is presented in Figure 2-2. The final Likert-type rating scale used in this evaluation tool has anchoring points with specific descriptors at 1, 3, and 5.

The Oxford Arthroplasty Technical Skill Score (OATSS) evaluation form

Surgical Step	1	2	3	4	5
Patient positioning	Unable to perform / Unsafe		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Skin incision and surgical approach	Unable to perform / Incorrectly placed incision / Unsafe approach		Performed with difficulty but no safety concerns		Accurate, safe & skillful approach
Joint inspection & assessment for contra-indications	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurate & skillful assessment
Osteophyte excision	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Seating of tibial saw guide onto tibia and estimation of resection height	Unable to perform / Over excision which compromised procedure		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Execution of the vertical cut for tibia	Unable to perform / Incorrect med-lat position / Unsafe (e.g. saw blade handle raised / plunged deeply risking posterior cortex damage)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Execution of the horizontal tibia cut and removal of tibial plateau	Unable to perform / Unsafe (e.g. MCL not protected)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Ensuring sufficient amount of tibial plateau is excised and further resection made if necessary	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed. Did not require "re-cut"
Insertion of the femoral IM rod with attention to patella	Unable to perform / Neglected to perform / Unsafe (e.g. caused damage to patella)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Assembly of the femoral drill guide & placement of the femoral drill holes	Unable to perform / Neglected to perform all alignment checks / Incorrectly performed		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Excision of the posterior femoral condyle with saw using saw guide	Unable to perform / Neglected to perform / Unsafe (e.g. saw blade plunged too deeply)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Milling and measurement of flexion / extension gaps	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Estimation of the trial tibial template	Unable to perform / Neglected to perform / Incorrect sizing with large medial overhang of template		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Preparation of the tibial plateau keel slots	Unable to perform / Neglected to perform / Unsafe clearing of keel slot		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Selection of meniscal bearing & ligament balancing	Unable to perform / Neglected to perform / Unsafe (e.g. risked stretching / avulsion of ligaments)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Cementation and/or implantation technique	Unable to perform / Unsafe (e.g. allowed cement to harden befor implantation was complete)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed. Limb in correct position during implantation
Attention to MCL / soft tissues	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		All carefully protected at all times.
Final inspection for bearing impingement and alignment	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed

Figure 2-2. The Oxford Arthroplasty Technical Skill Score (OATSS) evaluation form.

2.3. STUDY B: DETERMINING THE INTER-RATER RELIABILITY OF THE OXFORD ARTHROPLASTY TECHNICAL SKILL SCORE (OATSS)

2.3.1. Methods

2.3.1.1. Setting and participants

This prospective study took place over a four-month period in a large university teaching hospital (Nuffield Orthopaedic Centre, Oxford). Institutional approval was obtained and all participants recruited gave informed consent. The participants consisted of orthopaedic trainees (Arthroplasty Fellows and senior Specialty Registrars ST7-8) who were undertaking their training in the Arthroplasty Unit of the institution. All participants had surgical logbook evidence of having previously performed both TKAs and OUKAs (either independently or under supervision), indicating that they had the basic psychomotor skills to perform knee arthroplasty surgery. Trainees recruited into the study agreed to have their performance evaluated by two assessors whilst undertaking the OUKA procedure. All the procedures were supervised by experienced consultant knee surgeons who would give verbal instructions or take over where necessary.

2.3.1.2. Assessment of OUKA performance

Two assessors used the OATSS scoring system to evaluate 14 OUKA procedures. One of the assessors was the developer of the OATSS scoring system. The second assessor was the surgeon who was in charge of supervising the trainee during each procedure (i.e. one of four knee surgeons with international reputations as trainers of the OUKA procedure). These experts had been briefed about the use of the OATSS evaluation tool. Both assessors observed the trainee perform the entire

OUKA procedure (from patient set-up to wound closure). The two assessors then completed the OATSS forms independently. This consisted of rating the trainee's performance on each of the sub-steps of the OATSS form on the Likert-type rating scale, ranging from 1-5 (Figure 2-2). A trainee would be awarded only "1" point in instances where he or she did not know how to perform a particular step, or the supervising trainer had to take over. The maximum possible OATSS mark was 90 and the minimum mark was 18. Although the trainees were not informed of their OATSS marks, the supervising trainer did provide them with confidential feedback about their general performance during the procedure.

2.3.1.3. Statistical analysis

The inter-rater reliability of the overall OATSS score was assessed using the intraclass correlation coefficient (ICC), with ICC > 0.61 being considered substantial agreement. In addition, it was important to determine the inter-rater reliability for each of the 18 sub-sections of OATSS, to further demonstrate the reliability of each subsection category. This was assessed using the weighted kappa coefficient with $k > 0.61$ being considered substantial agreement[237]. All data were analysed with SPSS software.

2.3.2. Results

Table 2-2 summarises the overall OATSS marks awarded by the two assessors for the 14 OUKA procedures. The inter-rater reliability for OATSS marks was excellent (ICC = 0.96). Sub-analysis to take into consideration the subsections of OATSS demonstrated that the inter-rater reliability remained satisfactory (Weighted kappa coefficient, $k = 0.71$).

Oxford UKA procedure	Total OATSS rating (Assessor 1)	Total OATSS rating (Assessor 2)
1	49	50
2	69	72
3	61	64
4	72	76
5	57	54
6	59	66
7	76	77
8	56	54
9	50	44
10	67	67
11	63	59
12	73	72
13	69	71
14	72	73

Table 2-2 Summary of the overall OATSS ratings that the two assessors awarded to each of the OUKA procedures (maximum possible OATSS = 90).

2.4. STUDY C: DETERMINING THE CONSTRUCT VALIDITY OF THE OXFORD ARTHROPLASTY TECHNICAL SKILL SCORE (OATSS)

2.4.1. Methods

2.4.1.1. Setting and participants

A further prospective study was then conducted over a six-month period. The participants consisted of two cohorts: 'Expert' and 'Trainee'. The Expert cohort consisted of four consultant knee surgeons with a sub-specialist interest in OUKA surgery. The Trainee cohort consisted of five orthopaedic trainees (four Arthroplasty Fellows and one final year Registrar) who were undertaking their training in the Arthroplasty Unit of the institution. All participants had surgical logbook evidence of having previously performed both TKAs and OUKAs (either independently or under supervision). All the procedures that were performed by the Trainee cohort were supervised by experienced consultant knee surgeons who would provide verbal instructions or take over where necessary.

2.4.1.2. Assessment of OUKA performance

A trained assessor (AA) used the OATSS scoring system to evaluate 26 real-life OUKA procedures being performed by the two groups (13 by the Experts and 13 by the Trainees). This assessor had extensive experience in both the OUKA procedure and the use of technical skill scoring systems. In addition, the previous validation study (Study B) had demonstrated that this assessor had excellent agreement with expert OUKA surgeons with regards to skill evaluation using OATSS. As in Study B, the performances of the participants were rated using the OATSS scoring system. The OATSS evaluation forms were made freely available to all participants at the beginning of the study (and prior to each individual

procedure), but it was not possible for them to refer to this intra-operatively. This step was taken to ensure that the Trainees were familiar with the contents of OATSS and what was expected of them, since the Experts had some prior experience of using OATSS in previous pilot studies. The participants were not informed of their OATSS marks during the course of the study.

2.4.1.3. Statistical analysis

Descriptive statistics were carried out for all variables. The Shapiro-Wilk test showed that the data was normally distributed; therefore parametric tests were applied with data presented as means and standard deviations (SD). The independent group *t*-test was used to compare the technical performance of the cohorts based on their OATSS scores. All data were analysed with SPSS software. A p-value < 0.05 was considered statistically significant.

2.4.2. Results

The mean OATSS score for the Experts was 87.7 (SD 1.3) compared with a mean OATSS score of 63.2 (SD 8.7) for the Trainees, demonstrating the construct validity of OATSS for *in vivo* assessment of surgical performance during OUKA ($p < 0.0005$). Figure 2-3 illustrates the difference in performance between the two cohorts (as assessed by the OATSS evaluation tool).

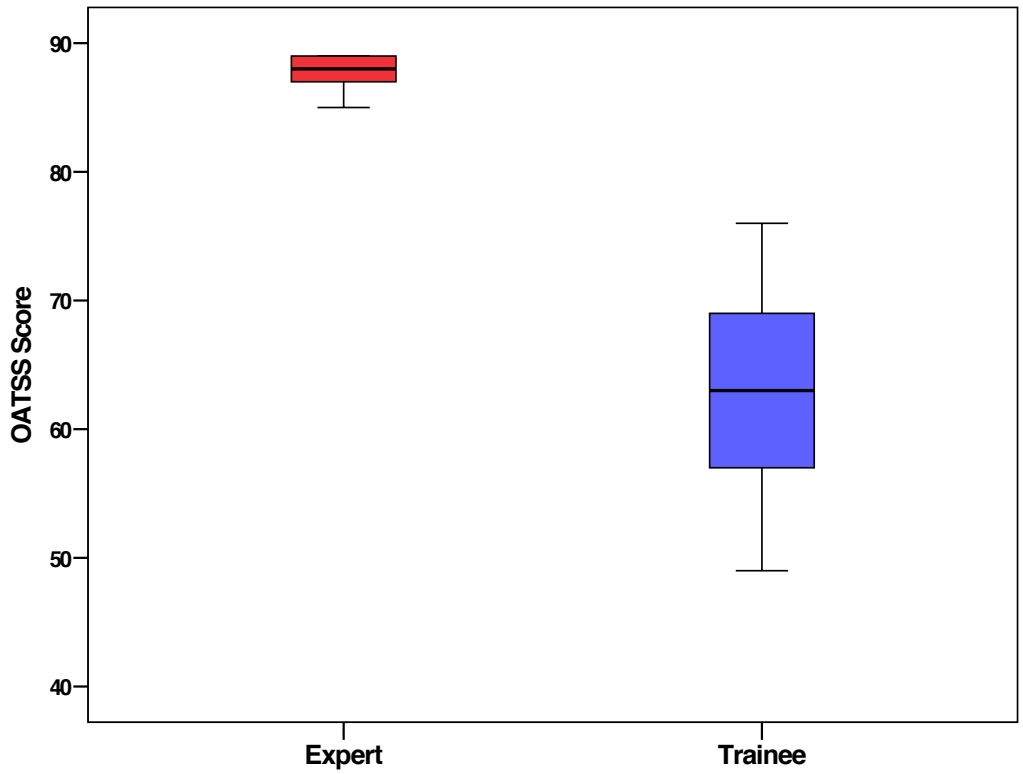

Figure 2-3 Box plots comparing the performance of the Expert (n=13) and Trainee (n=13) cohorts during Oxford unicompartmental knee arthroplasty. (OATSS = Oxford Arthroplasty Technical Skill Score. Maximum possible OATSS = 90. Box represents the interquartile range, bold line in the box represents the median, whiskers represent the 2.5 and 97.5 percentiles, and circles are outliers)

2.5. DISCUSSION

2.5.1. Implications of the studies

These three interlinked studies have enabled the development and validation of a procedure-specific assessment scale for the OUKA procedure. Study A used a consensus-based framework of expert opinion (Delphi methodology) to develop this assessment tool, namely the Oxford Arthroplasty Technical Skill Score (OATSS). Studies B and C subsequently demonstrated inter-rater reliability and construct validity of OATSS in the real operating theatre. This is the first time that a procedure-specific assessment scale has been developed and validated for an arthroplasty procedure.

The development and validation of procedure-specific assessment scales is an important aspect of surgical skills training and assessment. In addition to their use in assessment, these tools can be used to provide specific and meaningful feedback to trainees about their performance – a key component of the motor learning process[57,238]. Furthermore, they can be used to monitor a trainee's progress to determine when he or she has become proficient in a particular procedure. A number of procedure-specific assessment scales have been designed and validated by other surgical sub-specialities[213-216]. However, these have predominantly been developed through consultation with experts at a local (institutional) level rather than by obtaining expert consensus at an international level. Furthermore, the use of the Delphi methodology to develop technical skill assessment tools is a relatively novel technique. Peyre et al.[239] first used Delphi methodology to develop a task-specific checklist for laparoscopic Nissen fundoplication. Others have since gone on to develop and validate task-specific assessment scales for other laparoscopic procedures[227,229]. Within the field of orthopaedic surgical education, the Delphi technique has only been used for the development of task-specific checklists such as the OCAP PBAs[217], the Arthroscopic Surgical Skill Evaluation Tool (ASSET)[103] and the Imperial Global Arthroscopy Rating Scale (IGARS)[240]. The OATSS scoring system is the first consensus-based procedure-specific assessment scale in orthopaedic surgery that can be used as a stand-alone assessment tool for a complex orthopaedic procedure.

It is important justify the reasons for employing the Delphi technique for development of the assessment tool for OUKA as opposed to other techniques such as Cognitive Task Analysis (CTA) – another method which has previously been used in simulation-based surgical training[241-243]. CTA is an excellent means

of deconstructing an expert's knowledge of a surgical procedure into concrete measurable steps. It would therefore have been more appropriate for the development of a very detailed task-specific checklist. As previously mentioned, such checklists have been criticized for having low 'ceiling effects'[101]. The aim of this study was validate a reliable procedure-specific assessment scale that could be completed in a time-efficient manner by the assessor.

Use of the Delphi technique for the development of surgical skill assessment tools offers a number of other advantages. Firstly, conducting studies online meant that developers are not limited by geographical boundaries, and a large number of worldwide experts can be contacted over a short period of time in a cost-effective manner[220]. Secondly, the anonymous nature of this process means that influential experts will not be able to dominate and alter the opinion of others. There are also advantages when it comes to validity. The assessment systems generated using this approach will have high *content validity* (i.e. will have the essential elements of the skills that are being tested), due to the fact that they are developed using expert consensus opinion. The process also improves the *face validity* of a rating scale, because those who are regularly involved in teaching this procedure in the "real world" have highlighted the surgical steps.

When designing the questionnaire, several decisions were made to optimise the response rate and ease of completion of the form. A 3-point Likert scale was used, as it has been shown to be just as reliable and valid as a 5-point scale[244], whilst allowing the information regarding "ease of teaching the steps to trainees" to be included in the same one-page document. The response rate of 61% achieved in this study was consequently comparable with what is acceptable for survey response rates[245,246].

An interesting finding of this study was that only one Delphi round was necessary to reach expert consensus on the critical steps of OUKA, and that none of the steps in the original survey were discarded. This implies that there was very little disagreement between the experts with regards to the necessary surgical steps. A possible explanation for this finding is that the OUKA is a very *procedural* operation, which necessitates the exact execution of precise steps in a specific manner. This provides an exciting opportunity to devise a supplementary educational tool (in the form of a virtual reality simulator) for teaching the specific steps of the OUKA procedure (see Chapter 6).

As previously discussed, the reliability of a technical skill assessment tool is of paramount importance[92]. The inter-rater reliability of the total OATSS was excellent (ICC = 0.96). As sub-steps of the OATSS used a Likert scale to rate performance, the weighted kappa coefficient was also used to obtain a more meaningful insight into the reliability of OATSS. This again demonstrated good inter-rater reliability, with $k = 0.71$. It is possible that the Delphi methodology employed in this study helped to improve the inter-rater reliability of OATSS. Although evaluation of OATSS's *intra*-observer (test re-test) reliability would have been valuable, this was not logistically possible due to the difficulties of recording OUKAs for future playback. This would have necessitated the use of several mounted video cameras and microphones around the operating theatre on multiple occasions, and was not acceptable due to patient safety and manpower issues.

Previous studies that have validated Global Rating Scales have measured the *internal consistency* of the various "skill domains" (e.g. instrument handling and dissection) to ensure that they are consistent with one another[96]. This was not

appropriate in the case of OATSS, as this was a procedure-specific scale consisting of the sub-steps of the procedure rather than global skill domains.

The *construct validity* study demonstrated that the OATSS had excellent validity in the real operating theatre. However, the ceiling and floor effect of OATSS cannot be accurately determined by this study as highly skilled OUKA experts and orthopaedic trainees were compared. Future studies evaluating the psychometric properties of OATSS should include surgeons with wider-ranging skill level.

The issue of *feasibility* is also important as it can determine the widespread adoption of a technical skill scoring system. The OATSS appears to be a feasible assessment tool as it is a one-page document that has been successfully used in the real operating theatre. During previous pilot studies conducted by the author, it had become evident that one of the major criticisms of the current Procedure Based Assessments (PBA) is that they are excessively lengthy, and that assessors prefer more succinct assessment scales.

With further validation work and wider adoption, procedure-specific assessment tools such as OATSS could be used to monitor the learning curves of surgeons who are beginning to learn the OUKA procedure, to determine the point at which performance stabilises. This would, in turn, provide a guide as to the minimum number of supervised OUKA procedures that are necessary before an individual is safe to operate independently.

2.5.2. Limitations

There are limitations to the three studies described above:

- The psychometric properties of OATSS were not compared to any other validated technical skill assessment system. This was due to the fact that, at present, no such tool is available for the assessment of OUKA (or any other type of UKA). One possible option was modification of the OCAP PBA for TKR, but this is a task-specific checklist that has never been validated as a summative assessment tool. Furthermore, it would not have been possible to modify this for use in the assessment of OUKA due to the fundamental differences involved in performing these two procedures.

- The assessors in studies B and C were not 'blind'. It would have been more desirable (as is the case for many laparoscopic and arthroscopic skill validation studies), to have digitally recorded the entire procedure (with sound) and subsequently edited the videos to hide the identity of participants. However, this was not possible due to the logistics of recording a complex "open" procedure in a sterile and safe manner on multiple occasions. The author did conduct a pilot study, during which he recorded the procedures with a head-mounted camera whilst assisting the surgeons performing the OUKAs. Unfortunately, however, the quality of the video recordings (with the audio-visual equipment available at the time) was unsatisfactory due to movement artefacts and glare from theatre lights. Furthermore, some of the recordings were incomplete, as the surgeon performing the procedure would interfere with the "line of sight" of the assistant who was recording the OUKA procedure, resulting in the loss of critical assessment footage.

- It would have been desirable to have had more than one assessor for the construct validity study, but this was not possible due to the difficulty of finding a suitable second assessor who would be available at every OUKA procedure during the study period (six months).

2.5.3. Summary

These three interlinked studies have utilised the Delphi technique to develop an objective, procedure-specific assessment scale to assess technical skill during OUKA. In addition, they have demonstrated the reliability and validity of the resulting scoring system (Oxford Arthroplasty Technical Skill Score, OATSS) in the real operating theatre. Procedure-specific assessment scales have the potential to act as a powerful educational tool as they can provide meaningful and specific feedback on performance, in addition to determining when an individual has reached competency in a procedure. This is particularly important when evaluating technically demanding procedures that have a steep learning curve. Other possible surrogate markers to evaluate how well arthroplasty procedures have been performed, and the impact of these markers on clinical outcome, will be explored in the next chapter.

3. DEVELOPMENT AND EVALUATION OF A POST-OPERATIVE RADIOLOGICAL ASSESSMENT SYSTEM FOR UNICOMPARTMENTAL KNEE ARTHROPLASTY

3.1. INTRODUCTION

It is generally accepted that the most suitable method of assessing surgical performance is by direct observation of the procedure – either in real time or by viewing a previously recorded video. In reality, this is not always feasible so it would be desirable to identify indirect means of assessing surgical performance during a particular procedure. Although successful clinical outcome is the most important factor, it requires a long period of follow-up, especially in the case of arthroplasty procedures. An alternative, and traditionally used, method is to use post-operative radiological assessment as a surrogate marker of intra-operative performance. Despite some of the limitations of using this method (e.g. radiological assessment does not take into account the attention to soft tissue structures), surgeons often evaluate post-operative radiographs and pass some judgment as to how well the procedure was performed. Consequently, radiological assessment remains an important aspect of follow-up for knee arthroplasty procedures, and researchers continue to endeavour to improve the accuracy of this outcome measure.

In the case of TKA, the American 'Knee Society total knee arthroplasty roentgenographic evaluation and scoring system'[247] is a commonly used method of post-operative radiological assessment. This is because it assesses the parameters that are generally thought to be important in the successful outcome of TKA. For

example, it is well documented that good component alignment is necessary to ensure optimum clinical results following TKA[248-252]. However, there has been less focus on the relationship between radiographic parameters and clinical outcome following OUKA. The OUKA developers have recommended a number of radiographic parameters, including an arbitrary range for component positioning, as a guideline for post-operative radiological assessment[165]. Despite these guidelines, it is not clear what influence each of these radiographic parameters has on OUKA outcome. The influence of some of the parameters, such as tibial component positioning, has been investigated and suggests that deviation from the guideline values can result in poor clinical outcome[253]. Conversely, other studies have found that, in general, parameters such as component alignment have no significant influence on clinical outcome[254,255].

To date, no studies have attempted to incorporate all of the suggested radiographic parameters into a formal radiological assessment system that can predict clinical outcome. The development of such a tool would have several potential advantages. Firstly, it would enable surgeons to assess their own results and make appropriate adjustments to their surgical technique in order to achieve better results. Secondly, it would aid the identification of cases that have been carried out sub-optimally and are likely to result in early implant failure. Such cases could then be monitored more closely, enabling necessary interventions to be made in a timely manner. Finally, the assessment system could be used as a training tool to monitor the progress of trainees and provide them with constructive feedback on their performance.

3.1.1. Aims

The aim of this chapter was to develop and evaluate the utility of a post-operative radiological assessment system that correlates with clinical outcome following OUKA. This was conducted using two linked studies:

Study A: To use the Delphi methodology to determine important parameters for analysis during post-operative radiological assessment of OUKA.

Study B: To determine the value of post-operative radiographic parameters in predicting clinical outcome following OUKA.

3.2. STUDY A: Consensus-based Delphi Approach for Determining Important Parameters for Analysis During Post-Operative Radiological Asssessment of OUKA

3.2.1. Methods

3.2.1.1. Study design and participant selection

This prospective study employed the Delphi methodology (as in Study A of Chapter 2) to achieve consensus amongst an international group of expert knee surgeons on the important parameters that must be analysed during post-operative radiological assessment of OUKA. The participants were the same individuals who were contacted for Study A of Chapter 2.

3.2.1.2. Delphi design and determination of group consensus

The development of the Delphi survey began by outlining all the radiographic parameters that were highlighted by the developers of the OUKA prosthesis. In addition, other parameters identified in orthopaedic textbooks, peer reviewed

literature, or suggested by the Arthroplasty Unit faculty at the Nuffield Orthopaedic Centre (NOC) were also included. This information was used, in conjunction with guidelines on questionnaire design[231-234], to produce the preliminary survey. The resulting survey was piloted locally in order to ensure that it was user-friendly and unambiguous. As a result, some of the survey items were rephrased or clarified. The final survey (Figure 3-1) asked experts to indicate (on a 4-point Likert scale) the relative importance of each radiographic parameter in reflecting surgical performance and technique during OUKA (1 = Not important, 4 = Of vital importance). In addition, participants were encouraged to comment on any survey items, clarify their ratings, or add any "extra" parameters that they felt were particularly important.

The participants who agreed to take part in this study were asked to complete the Delphi survey and return it via email to the principle investigator. The intention was to repeat further Delphi rounds until group consensus was reached on the important post-operative radiographic parameters for OUKA assessment. In order to statistically quantify the reliability of the experts' opinions, Cronbach's α[235] was selected as a measure of internal consistency. An α value of 0.7 (i.e. agreement between 70% of experts) was selected to indicate group consensus, as this is an acceptable value of internal consistency for educational research purposes[236].

Component position on AP radiographs:

	Not important	Occasionally important	Important	Of vital importance
Femur: Varus / Valgus	☐	☐	☐	☐
Femur: Medial / lateral position	☐	☐	☐	☐
Tibia: Varus / Valgus	☐	☐	☐	☐
Tibia: Medial overhang	☐	☐	☐	☐
Tibia: Medial underhang	☐	☐	☐	☐
Bearing position (relative to tibial component)	☐	☐	☐	☐

Component position on Lateral radiographs:

	Not important	Occasionally important	Important	Of vital importance
Femur: Flexion/ Extension	☐	☐	☐	☐
Femur: Posterior overhang	☐	☐	☐	☐
Tibia: Posterior slope	☐	☐	☐	☐
Tibia: Anterior overhang	☐	☐	☐	☐
Tibia: Anterior underhang	☐	☐	☐	☐
Tibia: Posterior overhang	☐	☐	☐	☐
Tibia: Posterior underhang	☐	☐	☐	☐

Saw cuts:

	Not important	Occasionally important	Important	Of vital importance
Tibial spines undercut (during the horizontal cut)	☐	☐	☐	☐
Position of vertical tibial saw cut	☐	☐	☐	☐
Depth of vertical tibial saw cut	☐	☐	☐	☐
Depth of keel saw cut				
Breach in posterior cortex of tibia	☐	☐	☐	☐

Other parameters:

	Not important	Occasionally important	Important	Of vital importance
Lateral fit of component (gap between vertical wall and tibial spine)	☐	☐	☐	☐
Femoral cement in peg hole	☐	☐	☐	☐
Tibial cement mantle	☐	☐	☐	☐
Excess cement behind tibia / femur	☐	☐	☐	☐
Presence of posterior femoral osteophytes	☐	☐	☐	☐
Presence of excess bone or cement causing anterior impingement	☐	☐	☐	☐

Figure 3-1 Delphi survey to identify important parameters for analysis during post-operative radiographic assessment of OUKA.

3.2.1.3. Selection of radiographic parameters for assessment of OUKA

The radiographic parameters that 70% (or more) of the experts regarded as being "occasionally important", "important", or "of vital importance" for assessing surgical technique during OUKA were selected so that their influence on clinical outcome could be evaluated in the subsequent Study B.

3.2.1.4. Statistical analysis

Descriptive statistics in the form of means and standard deviations were calculated for each of the radiographic parameters being studied here. Cronbach's α was calculated for each radiographic parameter in order to determine the reliability of the experts' opinions. All data were analysed with SPSS software.

3.2.2. Results

A total of 56 experts were invited to participate and 30 took part in the survey resulting in a 54% response rate. Six were from the United Kingdom and 24 were from abroad.

Cronbach's α was greater than 0.8 for all the radiographic parameters in the survey indicating strong group consensus. It was, therefore, not necessary to repeat another Delphi round. None of the parameters were rated as "not important", thus all radiographic parameters that were present in the first Delphi round were selected for further evaluation in Study B. None of the survey participants suggested any additional radiographic parameters that should be assessed. The Cronbach's α values, Likert scale means and standard deviations for each radiographic parameter are summarised in Table 3-1.

Radiographic Parameter	Cronbach's α	Degree of importance (based on Likert scale)	
		Mean	Std. Deviation
Femur: Varus / Valgus	0.849	2.33	0.802
Femur: Medial / lateral position	0.847	2.93	0.740
Tibia: Varus / Valgus	0.853	3.00	0.910
Tibia: Medial overhang	0.863	2.73	0.740
Tibia: Medial underhang	0.859	2.70	0.877
Bearing position (relative to tibial component)	0.850	2.70	0.651
Femur: Flexion / Extension	0.850	2.60	0.855
Femur: Posterior overhang	0.849	2.27	0.907
Tibia: Posterior slope	0.862	2.67	0.758
Tibia: Anterior overhang	0.852	2.30	0.702
Tibia: Anterior underhang	0.848	1.83	0.747
Tibia: Posterior overhang	0.847	2.50	0.820
Tibia: Posterior underhang	0.850	2.63	0.964
Tibial spines undercut	0.856	2.33	0.802
Position of vertical tibial cut	0.854	3.00	0.743
Deep vertical tibial saw cut	0.852	3.13	0.860
Depth of keel saw cut	0.851	3.03	0.850
Breach in posterior cortex of tibia	0.898	3.81	0.688
Lateral fit of component	0.856	2.67	0.844
Femoral cement in peg hole	0.860	3.03	0.890
Tibial cement mantel	0.854	3.27	0.640
Excess cement behind tibia / femur	0.861	3.17	0.699
Presence of posterior femoral osteophytes	0.857	3.10	0.662
Presence of excess bone or cement causing anterior impingement	0.854	3.37	0.718

Table 3-1. Delphi survey results (Cronbach's α values and descriptive statistics) for identifying important parameters for analysis during post-operative radiological assessment of OUKA (n=32). (Likert scale: *1=Not important; 2=Occasionally important; 3=Important; 4=Of vital importance*)

The results of the Delphi survey were used to devise a list of important parameters for use during post-operative radiological assessment of OUKA (Figure 3-2). Detailed explanation of each of these parameters and their location on post-operative radiographs is provided in Table 3-2 and Figures 3-3, 3-4, 3-5.

Post-operative radiological assessment of Oxford unicompartmental knee arthroplasty

Radiographic Parameter	*Raw data*
Femur	
Varus / Valgus angle	
Flexion / Extension angle	
Medial / Lateral Position of implant	**Central / Off centre**
Posterior Fit (> 4mm overhang)	**No / Yes**
Anterior impingement of bone on bearing?	**No / Maybe / Yes**
Tibia	
Varus/Valgus angle	
Posterior tilt angle	
Medial fit (mm)	
Anterior fit (mm)	
Posterior fit (mm)	
Lateral fit	**Flush / Gap**
Meniscal bearing alignment	**OK / Misaligned**
Undercut tibial spines?	**No / Yes**
Vertical cut of tibia	**On medial tibial spine / Med to medial tibial spine / Into lateral tibial spine**
Depth of vertical cut	**OK / Deep**
Depth of keel cut	**OK / Deep**
Posterior cortex of tibia	**Intact / Not**
Implant-cement interfaces	
Femoral cement mantle	**Complete / Incomplete**
Tibial cement mantle	**Complete / Incomplete**
Other	
Posterior osteophytes	**No / Yes**
Cementophytes	**No / Yes**

Figure 3-2 Data collection form composed of important radiographic parameters used during post-operative radiological assessment of OUKA.

Radiographic Parameter	*Definition and method of assessment*	*Assessed on AP or LA view?*
Femur		
Varus / Valgus angle	Varus (+ve) or valgus (-ve) alignment of femoral component relative to the long axis of the tibia (with neutral taken as 0°)	AP (Figure 3.3)
Flexion / Extension angle	Flexion (+ve) or extension (-ve) of femoral component relative to long axis of posterior femoral cortex	LA (Figure 3.3)
Medial / Lateral Position of implant	Position of femoral component relative to the medial femoral condyle	AP (Figure 3.4)
Posterior Fit (> 4mm overhang)	Whether or not the femoral component was too large and overhung off the posterior aspect of the femoral condyle	LA (Figure 3.5)
Anterior impingement of bone on bearing?	Residual bone left on the anterior aspect of the femoral condyle impinging on the meniscal bearing during knee extension	LA (Figure 3.5)
Tibia		
Varus / Valgus angle	Varus (+ve) or valgus (-ve) alignment of tibial component relative to the long axis of the tibia	AP (Figure 3.3)
Posterior tilt angle	The posterior slope of the tibial component (7° posterior tilt is taken as neutral thus tilt >7° gives a +ve value and <7° gives –ve value)	LA (Figure 3.3)
Medial fit (mm)	Medial coverage of the tibial plateau by prosthesis (+ve values = overhang; -ve values = underhang)	AP (Figure 3.4)
Anterior fit (mm)	Anterior coverage of the tibial plateau by prosthesis (+ve values = overhang; -ve values = underhang)	LA (Figure 3.5)
Posterior fit (mm)	Posterior coverage of the tibial plateau by prosthesis (+ve values = overhang; -ve values = underhang)	LA (Figure 3.5)
Lateral fit	Gap between vertical wall of component and tibial spine	AP (Figure 3.4)
Meniscal bearing alignment	Bearing central and parallel to tibial component (indicated by the bearing shadow and x-ray marker within the bearing. Misaligned bearing defined as a bearing that is tilted or is more than 4mm medial from the vertical wall of tibial prosthesis)	AP (Figure 3.4)
Undercut tibial spines?	Horizontal saw cut causing undercutting of tibial spine	AP (Figure 3.4)
Vertical cut of tibia	Position of vertical saw cut relative to the tibial spine	AP (Figure 3.4)
Depth of vertical cut	Depth of vertical saw cut estimated by ingress of cement	AP (Figure 3.4)
Depth of keel cut	Depth of keel saw cut estimated by ingress of cement	AP (Figure 3.4)
Posterior cortex of tibia	Breach in posterior cortex of tibial plateau indicated by extrusion of cement posteriorly	LA (Figure 3.5)
Implant-cement interfaces		
Femoral cement mantle	Uniform cement penetration into the peg hole of the femur and beneath the femoral prosthesis	LA (Figure 3.5)
Tibial cement mantle	Uniform cement penetration beneath tibial prosthesis	AP (Figure 3.4)
Other		
Posterior osteophytes	Osteophytes present posteriorly on the femoral condyle	LA (Figure 3.5)
Cementophytes	Excess cement in the soft tissues around the prosthesis	AP (Figure 3.4) LA (Figure 3.5)

Table 3-2. Definitions of important post-operative radiographic parameters and how they are assessed.

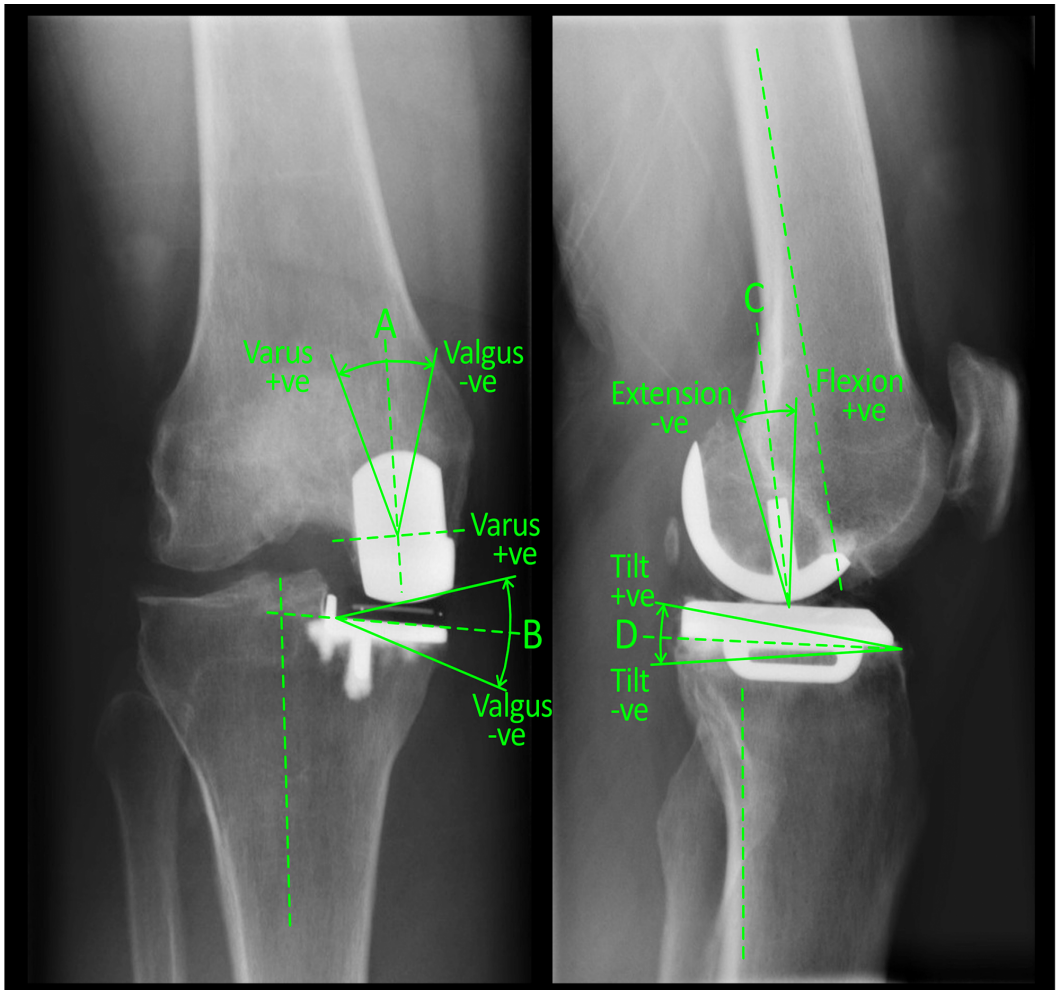

Figure 3-3 Post-operative AP and Lateral radiographs for assessment of component alignment, expressed as positive or negative degrees (A = Femoral component coronal alignment; B = Tibial component coronal alignment; C = Femoral component sagittal alignment; D = Tibial component posterior tilt).

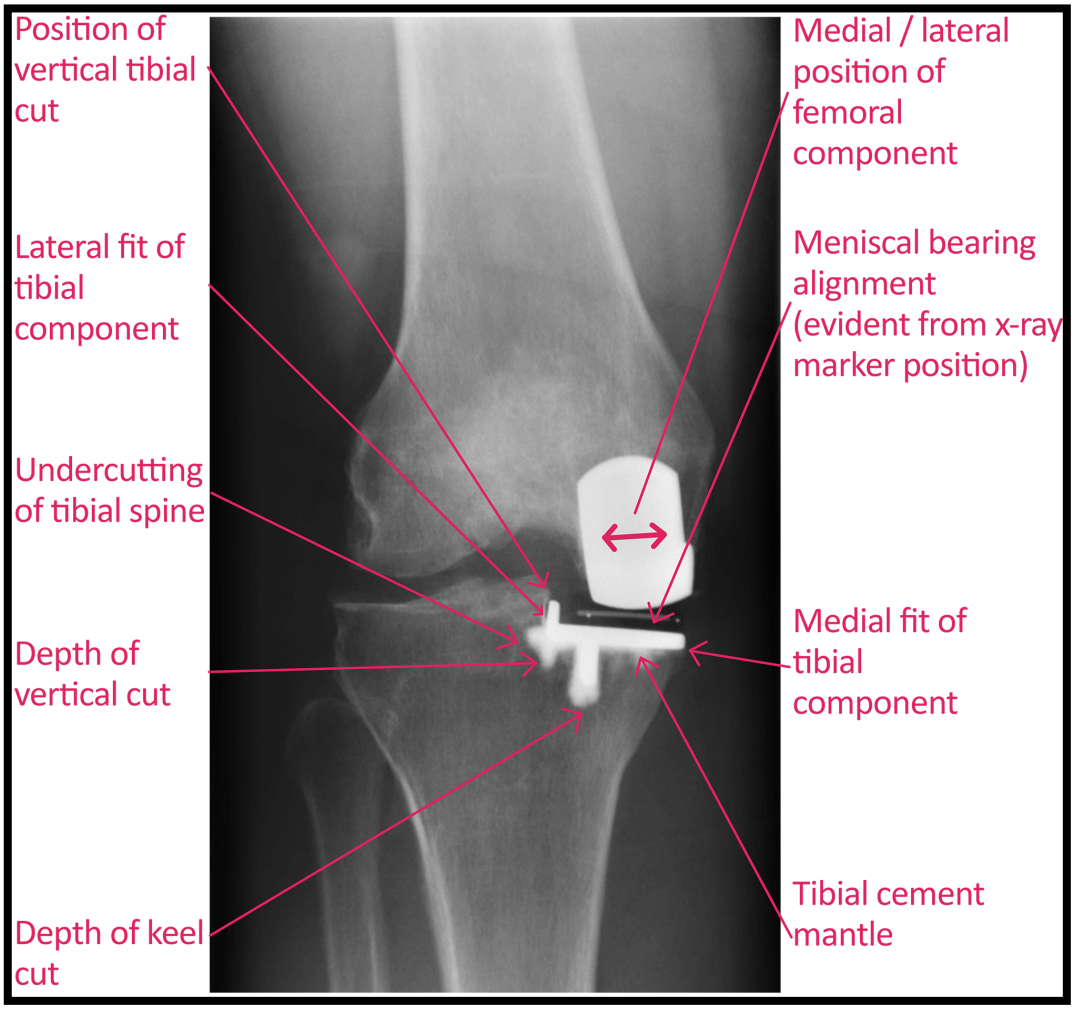

Figure 3-4 AP post-operative radiograph following medial cemented OUKA highlighting important parameters.

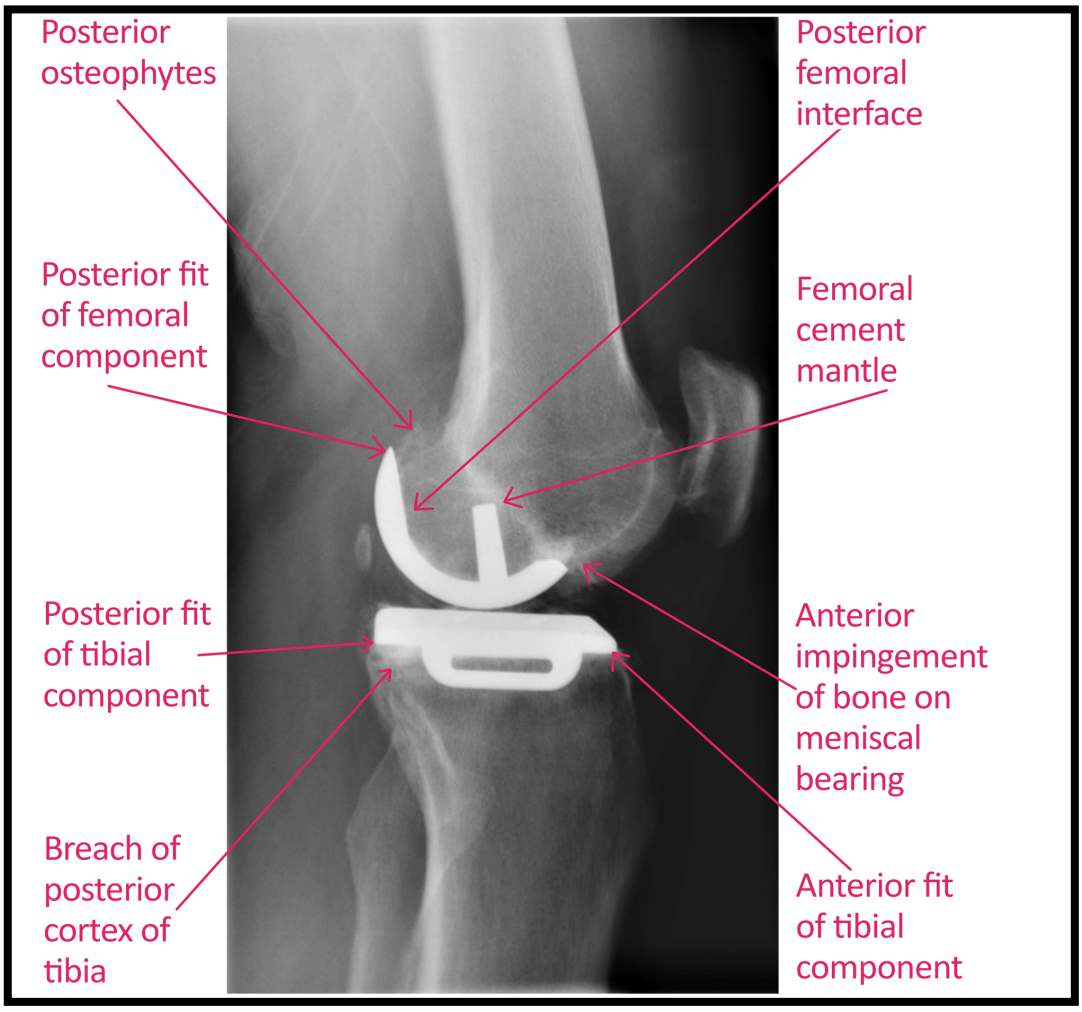

Figure 3-5 Lateral post-operative radiograph following medial cemented OUKA highlighting important parameters.

3.3. STUDY B: Determining the value of post-operative radiographic parameters in predicting clinical outcome following OUKA

3.3.1. Methods

For the purpose of this retrospective cohort study, clinical outcome following the OUKA procedure was defined as:

I. *Prosthesis failure* indicated by re-operation for any reason except infection.

II. *Functional outcome using the Oxford Knee Score (OKS)*[256]: This was expressed as the change between the pre- and post-operative OKS at one year (ΔOKS). The OKS at one year was selected as previous studies have shown that there is no statistical difference between scores at one year compared to those beyond this period[257]. For the purpose of this study (see below Section 3.3.1.1), patients with the best functional outcome were defined as those in the "Top Decile" for ΔOKS, whereas those with the worst functional outcome were defined as those in the "Bottom Decile" for ΔOKS.

3.3.1.1. Setting and patient cohorts

The NOC's OUKA database was used to identify 2000 patients who had undergone cemented medial OUKAs for anteromedial osteoarthritis between June 1998 and March 2009. The clinical indications for performing medial OUKA were the same as those previously described by its design group[165]. The aim was to analyse the post-operative radiographs of the relevant patients in order to determine which radiographic parameters influenced clinical outcome. The implication was that radiological assessment of these parameters could identify

"technical errors" made intra-operatively. As a result, patients who had undergone revision surgery for infection following their OUKA procedure were excluded, because it was not possible to prove that infection was due to a technical error.

The following sub-groups of patients were identified from the database for further analysis:

- ***Design Surgeon Failures*** = Patients with failed OUKAs who had originally undergone index surgery by the two surgeons responsible for developing the Phase 3 OUKA instrumentation and operative technique. These surgeons (DWM and CAFD) will be referred to as the OUKA Design Surgeons from now on.

- ***Design Surgeon Survivors*** = Patients who had originally undergone their index surgery by the OUKA Design Surgeons and who had not undergone revision surgery.

- ***Non-design Surgeon Failures*** = Patients with failed OUKAs had undergone index surgery by any surgeon other than the OUKA Design Surgeons. Consequently the surgeons in this group consisted of orthopaedic registrars, senior clinical fellows, and other arthroplasty consultants at the NOC.

- ***Non-design Surgeon Survivors*** = Patients who had not been revised and had originally undergone their index surgery by any surgeon other than the OUKA Design Surgeons.

- ***Design Surgeon Top Decile ΔOKS group*** = Top 10% of patients with the best functional outcome (determined by their ΔOKS) following surgery by the OUKA Design Surgeons.

- ***Design Surgeon Bottom Decile ΔOKS group*** = Bottom 10% of patients with the poorest functional outcome (determined by their ΔOKS) following surgery by the OUKA Design Surgeons.

Patients from the above sub-groups were then separated into three cohorts so that their post-operative radiographic parameters could be compared. Further details of these cohorts are presented below in Table 3-3. Thus, Cohort 1 compared OUKA *Design Surgeon Failures* with matched *Design Surgeon Survivors*, Cohort 2 compared OUKA *Non-Design Surgeon Failures* with matched *Non-Design Surgeon Survivors*, and Cohort 3 compared the *Design Surgeon Top Decile ΔOKS group* with the *Design Surgeon Bottom Decile ΔOKS group* (note that patients were not matched for gender or body mass index, as previous studies have shown that these factors do not have a significant effect on the outcome of UKA[258-261]).

Cohort	Sub-group X	Sub-group Z
Cohort 1 (Each patient in sub-group X was matched to two patients in sub-group Z based on "age", "surgeon", and "date of index operation")	***Design Surgeon Failures***	***Design Surgeon Survivors***
Cohort 2 (Each patient in sub-group X was matched to two patients in sub-group Z based on "age" and "date of index operation")	***Non-Design Surgeon Failures***	***Non-Design Surgeon Survivors***
Cohort 3 (All patients in this cohort had undergone surgery by the OUKA Design Surgeons)	***Design Surgeon Top Decile ΔOKS group***	***Design Surgeon Bottom Decile ΔOKS group***

Table 3-3 Details of the patient sub-groups in each cohort (note that comparisons were made between the radiographic parameters of each sub-group *within* a cohort, i.e. no comparisons were made *between* cohorts).

3.3.1.2. Radiological assessment

The post-operative radiographs consisted of antero-posterior and lateral views which were performed according to the standardised technique described by Goodfellow et al[165]. This technique ensures suitably aligned post-operative images that can be easily interpreted and used for comparison. The older radiographs, which were in photographic film format, were converted to JPEG files after being scanned digitally at 300 dpi resolution (Epson Expression 1640XL scanner with Transparency Unit, Epson, California, USA). The final analysis of all the radiographs was performed using custom image processing software written in

Matlab (v.7.0, The MathsWork Inc., Newton, MA, USA). The radiographic parameters identified by the Delphi procedure were analysed blindly by two trained reviewers.

3.3.1.3. Statistical analysis

Descriptive statistics were used to assess the baseline patient characteristics and radiographic parameters. The three cohorts were analysed independently, and in each one, the relative risk (RR) of *failure* or *poor clinical outcome* (i.e. low ΔOKS) for each radiographic parameter was calculated. For the purposes of performing relative risk analysis, the non-categorical radiographic parameters were sub-divided into 2 or 3 groups based on existing evidence in the literature for appropriate ranges for these values[253,254]. The ranges for these groups are listed in the RR summary tables in Section 3.3.2. The RR estimates are presented along with p values. The inter- and intra-observer reliability of the radiographic parameters was assessed using the intraclass correlation coefficient (ICC) for grouped continuous variables and weighted kappa for the categorical variables. Data were analysed in R statistical software (version 3.0.2; www.r-project.org) and SPSS software. A 5% significance level (i.e. $p < 0.05$) was considered significant throughout this analysis.

3.3.2. Results

Cohort 1: Comparison of OUKA Design Surgeon Failures and matched Design Surgeon Survivors

A total of 13 Design Surgeon Failures were identified and matched accordingly to 26 Design Surgeon Survivors. A summary of the comparison between the radiographic parameters for this cohort is presented in Appendix 1. The relative

risk ratio (for revision) of each radiographic parameter is presented in Table 3-4. Those with relative risk values higher than 1.5 are highlighted in the results table as these values were considered to have practical significance. The analysis showed that the only parameter that was associated with a statistically significant risk of failure was a misaligned bearing.

Cohort 1 Radiographic Parameter	A	B	Relative Risk	p value
Femur				
Varus / Valgus angle	>10°	-10° to +10°	Inestimable	-
Flexion / Extension angle	> 15°	0° to +15°	Inestimable	-
Medial / Lateral Position of implant	Off centre	Central	Inestimable	-
Posterior Fit (> 4mm overhang)	Yes	No	Inestimable	-
Anterior impingement of bone on bearing?	Yes	No / Maybe	Inestimable	-
Tibia				
Varus/Valgus angle	> -5° or + 5°	-5° to +5°	1.20	1
Posterior tilt angle	> -5° or + 5°	-5° to +5°	1.46	0.493
Medial fit (mm)	>3mm overhang	< 2mm overhang	1.75	0.572
Anterior fit (mm)	>3mm underhang	<2mm underhang	1.57	0.73
Posterior fit (mm)	>2mm overhang	< 2mm overhang	Inestimable	-
Lateral fit	Gap	Flush	0.457	0.107
Meniscal bearing alignment	Misaligned	OK	2.76**	0.047
Undercut tibial spines?	Yes	No	1.88	0.234
Vertical cut of tibia	Medial to medial tibial spine	On medial tibial spine	1.90	0.178
Depth of vertical cut	Deep	OK	1.41	0.617
Depth of keel cut	Deep	OK	Inestimable	-
Posterior cortex of tibia	Intact	Not	Inestimable	-
Implant-cement interfaces				
Femoral cement mantle	Incomplete	Complete	0.56	0.262
Tibial cement mantle	Incomplete	Complete	0.39	0.073
Other				
Posterior osteophytes	Yes	No	0.56	0.383
Cementophytes	Yes	No	0.78	0.724

Table 3-4. Influence of sub-optimal radiographic parameter values in column 'A' on risk of revision relative to values in column 'B' for Cohort 1 (e.g. a value of 2 implies that the risk of failure to those in category A is twice the risk of failure to those in category B). Instances where there are no cases in either column A or B are labelled as 'Inestimable'; ** indicates RR values that were statistically significant.

Cohort 2: Comparison of Non-Design Surgeon Failures and matched Non-Design Surgeon Survivors

A total of 34 Non-Design Surgeon Failures were identified and matched to 68 successful cases. A summary of the comparison between the radiographic parameters for this cohort is presented in Appendix 2. The relative risk ratio (for revision) of each radiographic parameter is presented in Table 3-5. The analysis showed that the only parameter that was associated with a statistically significant risk of failure was a misaligned bearing.

Cohort 3: Comparison of OUKA Design Surgeon 'Top Decile' ΔOKS group and the 'Bottom Decile' ΔOKS group

A total of 38 cases were identified that were in the Bottom Decile for ΔOKS and compared with 41 cases that were in the Top Decile for ΔOKS. Appendix 3 provides a summary of the radiographic parameters for this cohort. The relative risk ratio (for poor outcome) of each radiographic parameter is presented in Table 3-6. The analysis showed that none of the parameters were associated with a statistically significant risk of a poor functional outcome. The intra- and inter-observer reliability for radiological assessment was excellent, with ICC and weighted kappa values > 0.8 for all radiographic parameters.

Cohort 2 Radiographic Parameter	A	B	Relative Risk	p value
Femur				
Varus / Valgus angle	>10°	-10° to +10°	Inestimable	-
Flexion / Extension angle	> 15°	0° to +15°	Inestimable	-
Medial / Lateral Position of implant	Off centre	Central	Inestimable	-
Posterior Fit (> 4mm overhang)	Yes	No	Inestimable	-
Anterior impingement of bone on bearing?	Yes	No	4.58	0.232
Tibia				
Varus/Valgus angle	> -5° or + 5°	-5° to +5°	1.05	0.345
Posterior tilt angle	> -5° or + 5°	-5° to +5°	1.33	1
Medial fit (mm)	>3mm overhang	< 2mm overhang	0.62	1
Anterior fit (mm)	>3mm underhang	<2mm underhang	1.21	1
Posterior fit (mm)	>2mm overhang	< 2mm overhang	Inestimable	-
Lateral fit	Gap	Flush	0.60	0.059
Meniscal bearing alignment	Misaligned	OK	2.40**	0.007
Undercut tibial spines?	Yes	No	1.67	0.756
Vertical cut of tibia	Medial to medial tibial spine	On medial tibial spine	1.06	1
Depth of vertical cut	Deep	OK	0.26	1
Depth of keel cut	Deep	OK	Inestimable	-
Posterior cortex of tibia	Intact	Not	Inestimable	-
Implant-cement interfaces				
Femoral cement mantle	Incomplete	Complete	0.53	0.59
Tibial cement mantle	Incomplete	Complete	0.38	0.088
Other				
Posterior osteophytes	Yes	No	0.31	0.512
Cementophytes	Yes	No	0.34	0.066

Table 3-5. Influence of sub-optimal radiographic parameter values in column 'A' on risk of revision relative to values in column 'B' for Cohort 2 (e.g. a value of 2 implies that the risk of failure to those in category A is twice the risk of failure to those in category B). Instances where there are no cases in either column A or B are labelled as 'Inestimable'; ** indicates RR values that were statistically significant.

Cohort 3 Radiographic Parameter	A	B	Relative Risk	p value
Femur				
Varus / Valgus angle	>10°	-10° to +10°	Inestimable	-
Flexion / Extension angle	> 15°	0° to +15°	2.06	1
Medial / Lateral Position of implant	Off centre	Central	Inestimable	-
Posterior Fit (> 4mm overhang)	Yes	No	Inestimable	-
Anterior impingement of bone on bearing?	Yes	No / Maybe	2.07	0.237
Tibia				
Varus/Valgus angle	> -5° or + 5°	-5° to +5°	1.16	1
Posterior tilt angle	> -5° or + 5°	-5° to +5°	1.11	0.787
Medial fit (mm)	>3mm overhang	< 2mm overhang	2.18	0.334
Anterior fit (mm)	>3mm underhang	<2mm underhang	0.59	0.226
Posterior fit (mm)	>2mm overhang	< 2mm overhang	Inestimable	-
Lateral fit	Gap	Flush	0.59	0.063
Meniscal bearing alignment	Misaligned	OK	2.18	0.136
Undercut tibial spines?	Yes	No	2.30	0.271
Vertical cut of tibia	Medial to medial tibial spine	On medial tibial spine	1.28	0.512
Depth of vertical cut	Deep	OK	1.25	0.671
Depth of keel cut	Deep	OK	Inestimable	-
Posterior cortex of tibia	Intact	Not	Inestimable	-
Implant-cement interfaces				
Femoral cement mantle	Incomplete	Complete	0.76	0.317
Tibial cement mantle	Incomplete	Complete	0.57	0.053
Other				
Posterior osteophytes	Yes	No	0.90	0.794
Cementophytes	Yes	No	0.81	0.488

Table 3-6. Influence of sub-optimal radiographic parameter values in column 'A' on risk of revision relative to values in column 'B' for Cohort 2 (e.g. a value of 2 implies that the risk of failure to those in category A is twice the risk of failure to those in category B). Instances where there are no cases in either column A or B are labelled as 'Inestimable'; ** indicates RR values that were statistically significant.

3.4. DISCUSSION

3.4.1. Implications of the studies

These two studies were designed to identify important parameters for the post-operative radiological assessment of OUKA, and to determine if there was an overall relationship to clinical outcome. In Study A, the Delphi methodology demonstrated that there was a high level of consensus amongst experts as to what parameters are perceived to be important in post-operative radiological assessment for OUKA. Twenty-three radiographic parameters were highlighted as being good surrogate markers of how well the OUKA procedure had been performed. Study B subsequently attempted to determine whether clinical outcome could be predicted by post-operative radiological assessment of these parameters. The only parameter that appeared to be associated with an increased risk of prosthesis failure was a misaligned meniscal bearing i.e. a bearing that was tilted or was more than 4mm medial from the vertical wall of the tibial prosthesis. Evidence of bearing misalignment increased the risk of revision by 2.8 times for "design surgeons" and 2.4 times for "non-design surgeons". It was also 2.2 times more likely to result in a poor functional outcome. Several radiographic parameters were associated with a poor clinical outcome but none of these associations reached statistical significance. Thus, it was not possible to devise a unifying "radiological scoring system" that could predict clinical outcome of OUKA.

The value of decision making by consensus, especially when dealing with complex clinical situations, is now being highlighted more frequently by the academic orthopaedic community[262]. The advantages of using the Delphi technique to determine expert consensus in an area of relative uncertainty have

already been discussed in Chapter 3. These include high content validity, cost-effectiveness, and capture of a wide geographical area. Based on the currently available literature, this appears to be the first study that has used the Delphi methodology to determine the important radiographic parameters for post-operative assessment of OUKA. The high internal consistency of the survey responses (Cronbach's $\alpha > 0.8$ for all questions) demonstrates that there was a high level of agreement amongst the experts as to what parameters are important for assessment. This is further supported by the fact that consensus was reached after the first Delphi round. Consequently, a comprehensive list of radiographic parameters was compiled for post-operative radiological assessment of OUKA that would be of educational value for feedback purposes. However, it was not possible to use the parameters (based on their relative risk ratios) to develop a "radiological scoring system" that could predict the risk of prosthesis failure or poor clinical outcome. The only parameter that was associated with an increased risk of prosthesis failure was a misaligned meniscal bearing. A number of factors (which are commonly interlinked) can contribute to bearing misalignment. These include impingement from soft tissues, cement, or osteophytes. This problems caused by bearing misalignment has previously been alluded to by other investigators, who suggested that it could increase the risk of impingement, pain, and dislocation (as a result of the bearing spinning)[165,263]. Placing the femoral peg drill holes (and hence the femoral component) too medially can result in excessive medialisation of the bearing and increase its distance from the vertical wall of the tibial component. This can in turn allow the bearing to spin, maltrack, and even dislocate. Despite the possibility of this technical error, existing literature suggests that component alignment has very little influence on the outcome of the mobile

bearing OUKA[255], since it is relatively easy to implant the prosthesis in the range recommended by the developers[254,264].

No other studies have previously been able to devise a radiological scoring system that can predict clinical outcome of UKA, and it remains to be seen whether or not this can be achieved. One study attempted to do this by giving arbitrary points to the radiographic parameters recommended by the OUKA developers and determining the correlation of this score with Oxford Knee Scores (OKS) (at a mean of 4.5 years)[265]. They found no correlation between this radiographic score and the OKS. However, this study was limited by its small sample size, the manner in which the radiographic score was devised, and the lack of standardised post-operative radiographs.

One possible explanation for the lack of statistically significant associations between radiographic parameters and clinical outcome may be the small sample sizes used, or the fact that there were too few cases in some of the relative risk categories to make appropriate estimations. For example, it has been well documented that a periprosthetic tibial plateau fracture is a complication that results from errors in surgical technique during UKA procedures[266-268]. However, since there were no cases of a tibial plateau fracture in the "design surgeon" or "non-design surgeon" failures, it was not possible to determine the link between this complication and a relevant radiographic parameter such as a breach in the posterior cortex of the tibia. This example illustrates that one must bear in mind that the lack of correlation between the radiographic parameters and clinical outcome in the current study does not necessarily imply that the parameter is not important. For example, one of the parameters that had a high relative risk ratio for revision, but was not statistically significant, was anterior boney impingement on

the meniscal bearing. Anterior impingement is caused by excess bone on the anterior aspect of the native femoral condyle, which then impinges on the anterior aspect of the bearing during knee extension. Previous studies analysing retrieved meniscal bearings have demonstrated that bearing impingement can result in accelerated wear of the polyethylene bearing and cause implant failure[269,270]. Similarly, although increased risk of poor outcome from sawing errors was not observed in the current study, previous investigators have reported complications that can result from these surgical errors[271-274].

3.4.2. Limitations

There were some limitations in these two studies. Firstly, the sample sizes in the cohorts were small, which limited some of the relative risk ratio analyses. Unfortunately, control over this was not possible as the study was performed at a single institution. However, the reason for the small sample size was not due to the size of the institution's patient pool. In fact, the "design surgeon failures" were selected from a pool of over 1000 patients who had undergone surgery by the two design surgeons. Similarly, the "non-design surgeon cohort" was selected from a pool of over 1000 patients who had undergone surgery by non-design surgeons. The limiting factor was that there were very few failures in either cohort. Future multicenter studies could analyse these radiological parameters in larger cohorts of patients with a higher proportion of "failures" in order to further evaluate the role of radiological parameters.

3.4.3. Conclusions

The results of this study have identified what are perceived by experts to be the most important radiographic parameters that should be assessed following OUKA,

and allow for the publication of an x-ray scoring system. This set of parameters could be used to provide surgeons (especially trainees) with systematic feedback about specific aspects of their surgical technique and improve their training. Based on the findings of the current study, patients whose post-operative radiographs highlight meniscal bearing misalignment should be followed up with greater vigilance as they may be at greater risk of early prosthesis failure.

4. DEVELOPMENT AND VALIDATION OF A COGNITIVE SIMULATOR FOR TEACHING THE OXFORD UNICOMPARTMENTAL KNEE ARTHROPLASTY PROCEDURE

4.1. INTRODUCTION

The field of surgical education and training has attracted considerable attention over the past two decades due to a combination of factors. These include the evolution of surgical technology and the accompanying growth of new surgical procedures (in particular minimally invasive surgery), the reduction in surgical trainees' working hours[275], and the increasing economic pressures on healthcare systems. Furthermore, a focus on the improvement of patient safety and the reduction of surgical errors means that the traditional apprenticeship model of surgical training, whereby the only way of gaining surgical experience was to practice on patients (albeit under supervision), is now being challenged. Consequently, there has been an increased focus on methods to improve surgical training in the face of these new challenges[90,276-278].

One of the most important developments has been the introduction of *simulation-based education*[279], which provides a controlled setting in which to teach and assess a wide range of skills without risk to patients. There is now some promising evidence that demonstrates the transfer validity of simulation training to the real operating theatre in a number of surgical disciplines[137,146,280,281]. Advances in computer technology have meant that Virtual Reality (VR) simulators can now represent a wide range of surgical procedures. However the key issue is not the degree of fidelity (realism), but whether or not a simulator is effective at teaching

or assessing a particular skill. This means that the simulator's content and the manner in which the simulation-based education is delivered is more important than the actual complexity of the technology. The science of learning theory and human performance has consequently been applied to optimise the structure, content, and delivery of simulation-based surgical training[282].

4.1.1. The role of non-technical skills in surgery

Although technical ability is undoubtedly a key attribute of a competent surgeon, non-technical skills are also essential in the development of surgical expertise.[283,284] These can be broadly categorised under cognitive factors (which include knowledge and decision-making abilities), and personality traits (which include communication skills) with studies demonstrating that underperformance in these domains contributes to poor clinical outcome[285]. Consequently, those responsible for designing and delivering simulation-based education have made great efforts to develop simulators that can teach non-technical as well as technical skills, with studies (outside of orthopaedic surgery) indicating that VR simulators can be used to improve pre-operative planning, cognitive knowledge, intra-operative decision-making and situational awareness[144].

4.1.2. Are non-technical skills important in UKA surgery?

It is well recognised that OUKA has a steep learning curve[185], and this may in turn impact on the clinical outcome of the procedure. This is further exacerbated by the fact that most surgeons learn to perform OUKA at a later stage in their training. Therefore, it is imperative that every effort is made to improve training methods during this initial learning curve period, as they can have a positive impact on surgical performance. Simulation-based education using VR to improve cognitive

skills is an attractive option for bridging this gap as it enables trainees to focus on improving other areas, such as the technical aspects of the procedure. The educational value of laparoscopic VR simulators in improving the cognitive skills of junior surgical trainees has already been demonstrated[286,287], but no such studies exist in orthopaedic surgery. It has now become evident that traditional models of surgical education are no longer the most suitable training methods for modern-day surgical trainees, and that simulation-based education must play a major role in filling this shortfall. Trainees must enhance their cognitive (as well as technical) skills away from the real operating room in an educationally sound environment without risk to patients. For this to take place effectively, it is important to ensure that the cognitive simulators developed for this purpose are easily accessible, simple to use, possess content validity, and can demonstrate a significant educational impact.

4.1.3. Aims:

The aim of this chapter was to develop and validate a VR-based simulator for teaching the cognitive skills necessary to perform the OUKA procedure.

4.2. METHODS

4.2.1. Development of the cognitive simulator

The cognitive simulator, which has been named 'Virtual Reality Tool for Arthroplasty Learning' (ViTAL), is a software-based learning environment that aims to teach the cognitive skills necessary to perform the OUKA procedure. It was developed over a twelve-month period by the primary author through collaboration with a computer scientist and a graphic artist. The primary author

was responsible for development of the 'script board' of educational material and outlined the way in which this should be relayed to learners. This was then transferred to the ViTAL platform using the expertise of the graphic artist and computer scientist.

The ViTAL software runs on a standard desktop computer, and the development framework is composed of a multiplatform core engine and an abstract framework written using C and C++ programming languages. Authenticity of the learning software was enhanced by incorporation of patient-specific anatomical geometry extracted from MRI scans, along with utilisation of the CAD (Computer-Assisted Design) models of surgical instruments used in OUKA. The software framework for ViTAL was constructed on an open source rendering engine called Object-Oriented Graphics Rendering Engine (OGRE)[288]. This engine provided the visualisation capability for patient anatomy and surgical instruments. In addition to this core layer, the ViTAL framework enables user interactions with multiple input devices such as a keyboard, computer mouse, and more specialised devices such as a haptic stylus for force feedback (Note that although this version of ViTAL did not utilise haptic software, it was felt that the capability to support this technology should be built into the system from the outset for future use). Various training scenarios can be constructed using a purposely-designed programming language known as ViTAL Markup Language. Due to the simplicity of this language, it is possible to create the procedural elements of OUKA without in-depth knowledge of computer programming.

The primary purpose of ViTAL is to train and develop a surgeon's cognitive skills. The simulator uses an interactive "point and click" user-interface that is accompanied by a voice-over and textual information (Figure 4-1). The user is

guided through the procedure in a step-by-step manner whilst being prompted by icons and descriptive text (Figure 4-2). In order to successfully progress through the simulated OUKA procedure, objectives for each surgical step must be correctly executed. Surgical instruments are visually represented as both 3D graphics and 2D icons that must be selected by the user in order to progress through the OUKA procedure. When incorrect actions are performed, either hints are made or an animation will be triggered in the virtual operating theatre to demonstrate the correct execution of a particular surgical step on the virtual patient. In addition to teaching the surgical steps of the OUKA procedure, ViTAL focuses on topics such as operative indications, pre-operative radiographic assessment, descriptions of OUKA surgical instruments, and key intra-operative decision-making skills.

Rigorous test runs were performed on the simulator by the primary author and volunteers to ensure the learning material was conveyed clearly and that the user-interface was intuitive. Furthermore, throughout the development of ViTAL, there was consultation with a surgical team consisting of two surgeons involved in the design and development of the OUKA prosthesis, and two expert UKA surgeons who are instructors on the "Oxford Partial Knee Advanced Instructional Course".

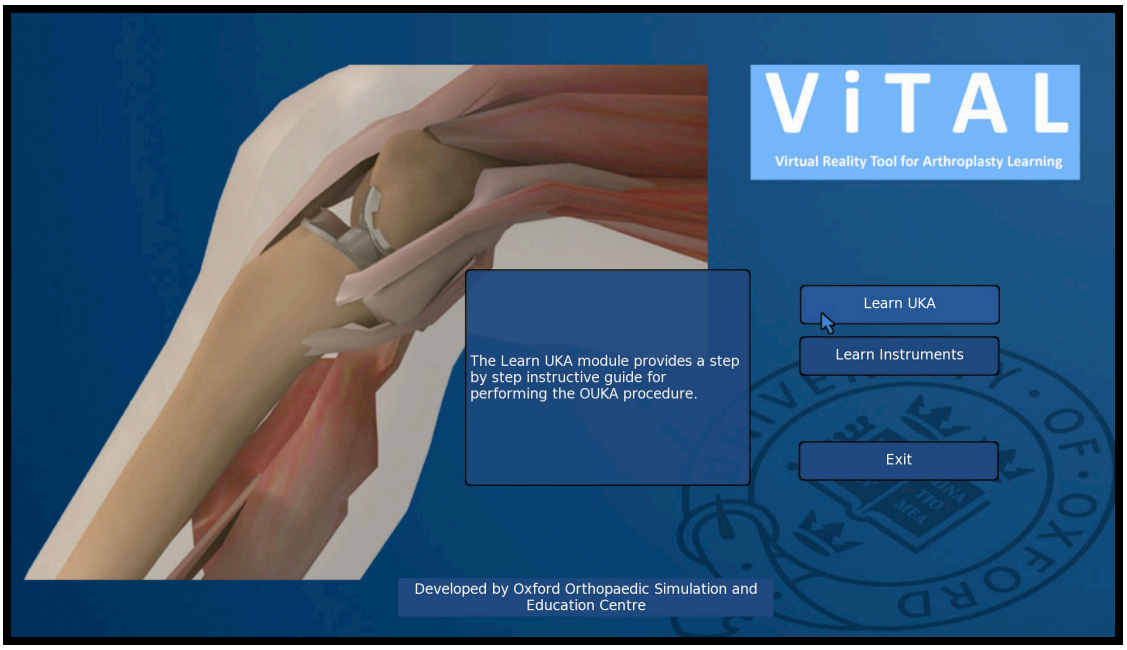

Figure 4-1: Screenshot of the ViTAL cognitive simulator user-interface demonstrating the "Home Screen".

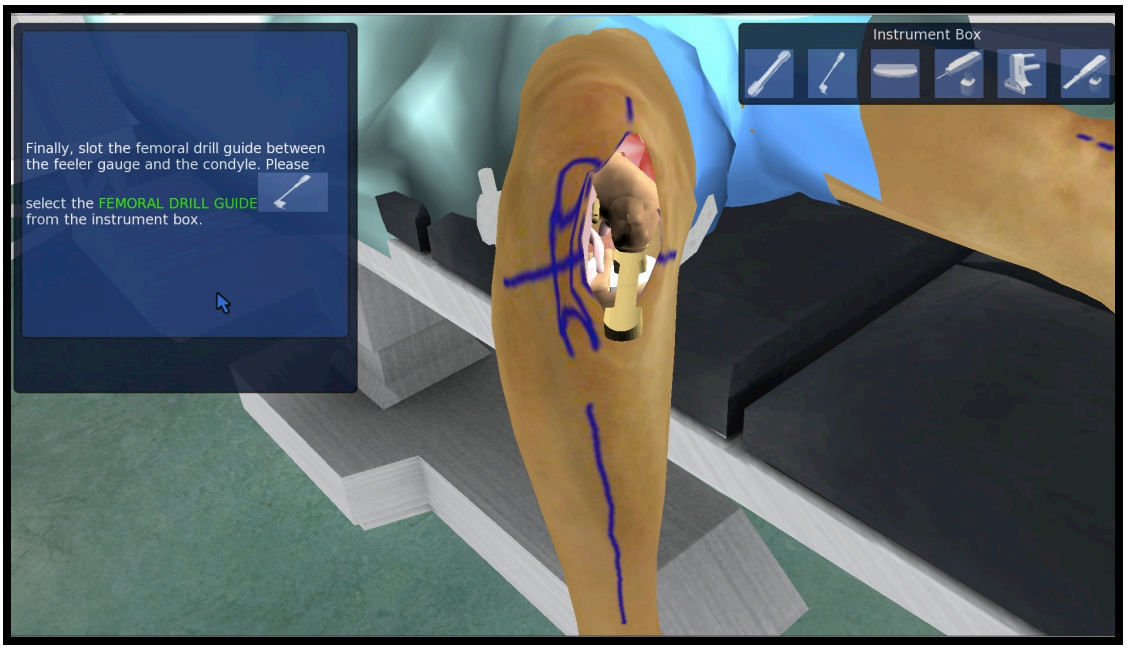

Figure 4-2: Screen shot of the ViTAL cognitive simulator user-interface demonstrating the "femoral drill guide assembly".

4.2.1.1. Determining the content of the ViTAL simulator

The main framework for the content of the simulator was based on advice from the surgeons who had designed the surgical procedure, or who taught on international instructional courses for OUKA. The aim was to convey the contents of the OUKA procedure's instructional manual and DVD in a more user-friendly and interactive way. However, in order to further optimise the educational potential of ViTAL, opinion was sought from leading OUKA experts from around the world in order to determine which particular aspects of the procedure are most difficult to teach orthopaedic trainees. This survey was conducted at the same time as the Delphi survey (detailed in Chapter 2), which had aimed to determine the key surgical steps of the OKA procedure. Expert OUKA surgeons in Europe, North America and Australasia were invited to participate in the survey via email. In addition to participating in the aforementioned Delphi process, each expert was also asked to indicate whether or not they found each of the surgical steps easy or difficult to teach their orthopaedic trainees (Figure 2-1, Chapter 2). The aim was to highlight aspects of the OUKA procedure that are difficult to teach so that further emphasis could be placed on them during the development of ViTAL.

4.2.2. Usability study to improve ViTAL

Software-based educational tools can be categorised under "Visual Display Terminals", and their usability is described as *"the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments"*[100]. When discussing the usability of a software system being used in medical informatics, the *context of use* by the user groups is of paramount importance[289]. Thus in the case of ViTAL, one must take into account the *user group* (surgical trainees or operating room staff), the *goal* (teaching cognitive

skills required to perform a surgical procedure), and the *working environment* (self-directed learning outside the real operating room). A usability study was therefore undertaken to qualitatively evaluate the perceived effectiveness of the ViTAL software and improve its usability.

Institutional approval was obtained and all participants gave informed consent. In total, 33 participants took part in this usability study. Individuals with a wide-ranging background were selected, including expert UKA surgeons, orthopaedic higher surgical trainees, junior doctors, operating room nurses, fifth year medical students, and biomedical scientists. Each participant was invited to explore the ViTAL software at their own pace under the supervision of the primary author. Participants undertook part of this process by viewing the procedure in stereoscopic 3D. Having comprehensively reviewed all aspects of ViTAL, the participants completed the usability questionnaire.

The purpose of first part of the questionnaire was to obtain user demographic information – which included level of computing knowledge, frequency/type of video game use, and previous exposure to UKA surgery (Appendix 4). The second part of the questionnaire (Figure 4-3) consisted of questions designed to gauge user opinion regarding software interface elements such graphics, textual information, accessibility to the simulator, and its use in orthopaedic training. These questions were based on a previous study exploring the opinions of orthopaedic surgeons on VR simulators[290]. In addition, to the general usability questions, clinicians and operating room staff were asked to comment on the face and content validity of the ViTAL cognitive simulator and its perceived efficacy as a training tool for the OUKA procedure. A free text box was also available for additional comments. Participants used a 5-point Likert scale (ranging from 1 =

strongly disagree to 5 = strongly agree) to provide feedback on the usability of the simulator. The results of this usability study were subsequently used to make changes to further improve the cognitive simulator. (These results are summarised in Table 4-3 section 4.3.2.)

Usability Survey *(for completion by all subjects)**:***

Software Usability Questions	Strongly Disagree	Disagree	Indifferent	Agree	Strongly Agree
Software interface was too complex					
The software is slow / unresponsive to your actions					
Textual information was not sufficient					
It was difficult to associate the textual description of the procedure with the graphics					
Additional "voice-over" information would make the instructions easier to follow					
The procedure should be displayed/viewed in stereoscopic 3D					
This software should be accessible on line for remote use 24 hours a day					
I found it enjoyable to learn using the software					

Usability Survey *(for completion by Clinicians and OR Staff only)**:***

Software Usability Questions	Strongly Disagree	Disagree	Indifferent	Agree	Strongly Agree
The virtual representation is a realistic view of procedure					
The tasks reflect those encountered in real-life surgery					
I would use this software as a supplementary tool for learning / teaching the Oxford UKA procedure					
The software teaches the procedural steps you need to perform the Oxford UKA					

Figure 4-3. Questionnaire used in the usability study to improve the ViTAL cognitive simulator.

4.2.3. Randomised controlled trial to determine the effectiveness of ViTAL

This randomised controlled trial took place in the Oxford Orthopaedic Simulation and Education Centre over a three-month period. Institutional approval was obtained and all participants gave informed consent. Twenty-five medical students, who were in the fifth year of their medical school training, were recruited during their rotation through the institution's orthopaedic department. As a part of their normal medical school training, all students undertake a compulsory orthopaedic attachment. Participation in this study was, however, voluntary. All participants completed a questionnaire to determine basic demographic data, their level of computing knowledge, frequency of playing computer games, and intended future career choice. Using sealed, opaque and consecutively numbered envelopes, the participants were randomised to receive either one of two methods of cognitive skills training: 'Standard Training' (control group) or 'ViTAL Training' (intervention group). Figure 4-4 illustrates the flow of participants through the trial.

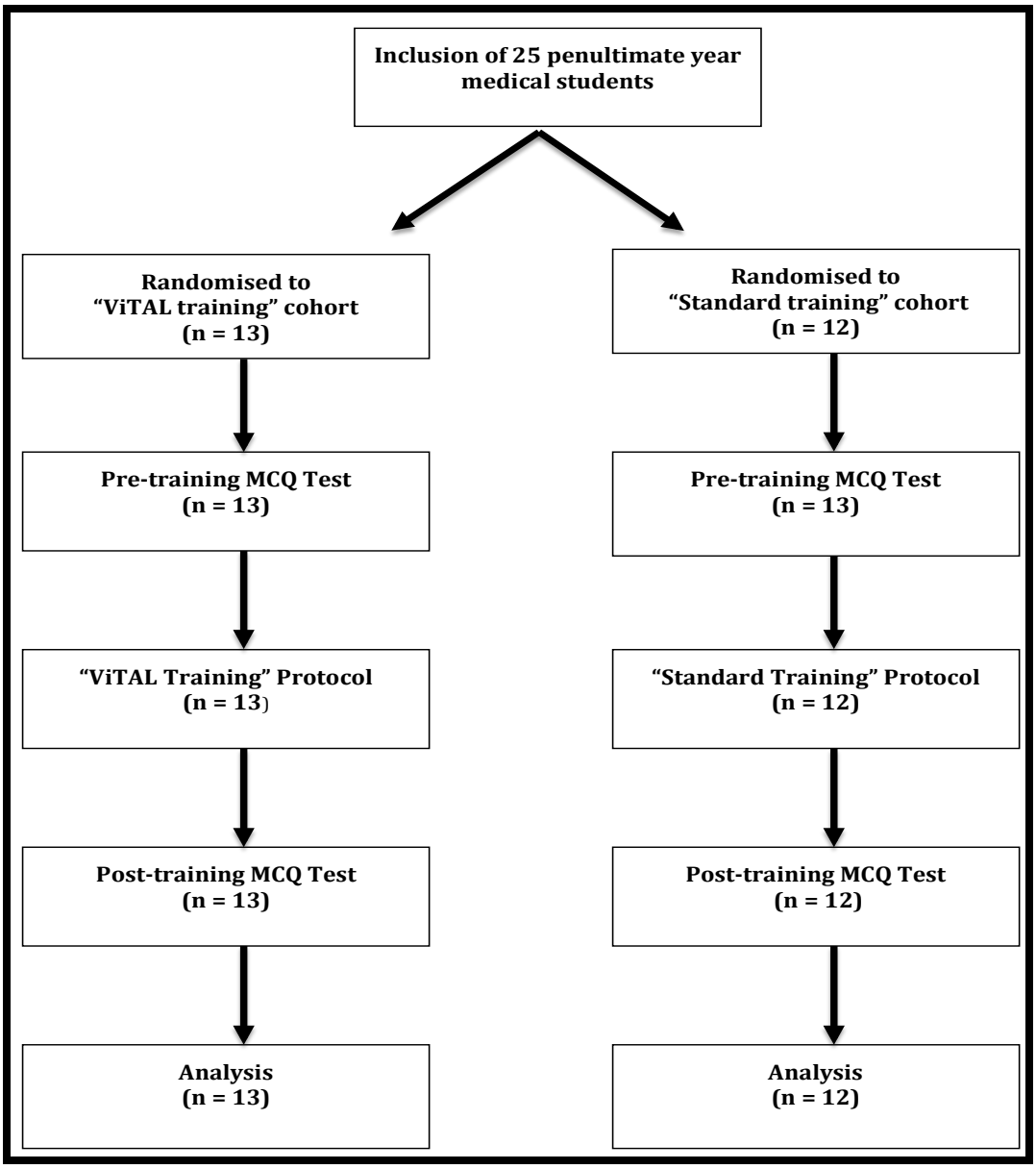

Figure 4-4: Flow diagram of participants through trial.

4.2.3.1. Training protocol

The Standard Training cohort (n = 12) used the OUKA surgical technique manual and operative procedure DVD (available from Biomet Inc., Warsaw, Indiana, USA) to learn about the OUKA. This surgical technique manual and DVD are well-established educational resources that were developed by the designers of the procedure. They are the most frequently used educational resources for orthopaedic trainees to learn the procedural steps of this operation (in addition to

attendance at an instructional course and "on the job" training in the operating room). The ViTAL Training cohort (n = 13) used the ViTAL simulator *in addition* to the surgical technique manual and DVD. The cognitive skills training for each participant took place over a three-week period. The frequency and duration of the training period was dictated by three factors. Firstly, the academic commitments of the medical students during the study period limited more frequent visits. Secondly, extending the training period beyond three weeks would not leave enough time in the study timetable to recruit, train and assess all the participants. Finally, the longer the study continued, the greater the probability of *contamination* of the training methods due to some participants being exposed to real-life OUKA cases, thereby placing them at an unfair advantage. All participants were required to read the surgical technique manual in their own time during the course of the trial (at least once a week). The Standard Training cohort attended the department's simulation centre once a week to watch the operative procedure DVD under the supervision of the primary investigator. However, no additional teaching was provided to the students during these visits. The ViTAL Training cohort also attended the simulation centre on a weekly basis to watch the operative procedure DVD. In addition, this group used the ViTAL simulator during the same visit. On the first visit of the ViTAL Training cohort, the concept of the cognitive simulator and its proposed educational objectives were introduced during a standardised 5-minute briefing period. This involved a demonstration of how to use ViTAL but no specific teaching with regards to the OUKA procedure. The two cohorts did not receive any further teaching on the use of ViTAL or the OUKA procedure during subsequent training visits.

4.2.3.2. Assessment of cognitive skills

Baseline levels of cognitive skills relating to the OUKA procedure were compared on the first visit using a 'pre-training test' consisting of twenty multiple-choice questions (MCQs). The effectiveness of the two different training methods was assessed using a 'post-training test' that consisted of a different set of MCQs to those used in the pre-training test. This test was conducted within one day of the third (final) visit to the simulation centre under the supervision of the primary investigator. The MCQs were "single correct answer" type with four options for each stem (Appendix 6). The questions were designed to test the modern-day indications, procedural knowledge, and key decision-making skills involved in OUKA. The relevant answers could be found in all three learning resources (surgical technique manual, DVD, and ViTAL). Some of the test questions are currently used in the established "Oxford Partial Knee Advanced Instructional Course". The remaining questions were written by the primary author and an expert OUKA surgeon, who had been involved in the development of the OUKA surgical technique manual and DVD. At the time of writing the test questions, the educational material covered by the ViTAL simulator was deliberately withheld from the expert OUKA surgeon, in order to ensure that the questions were not designed in a way that would favour the ViTAL Training cohort.

4.2.3.3. Statistical analysis

Target sample size:

Determining the effect size of a new cognitive simulator can be difficult as variability depends on a number of factors that include simulator characteristics, practice schedules, and participants' experience levels. There was no pilot data to

determine the educational effect of ViTAL in a medical student population. However, education research in the field of psychology suggests that large effect sizes (Cohen's *d* values greater than 0.8) are considered acceptable for a teaching intervention[291]. Furthermore, a recent systematic review and meta-analysis on technology-enhanced simulation for healthcare education discovered a pooled effect size of 1.2 for knowledge outcomes following simulator-based training[292]. Thus to detect an effect size of 1.2, using a two-tailed alpha of 0.05 and a power of 0.8, it was necessary to recruit at least 12 participants in each cohort. Fifteen participants were therefore recruited in each cohort. Summary statistics are provided for the ViTAL development survey and usability study.

There were two main outcome variables:

1) *The generic effect of training on cognitive skills acquisition.* This was measured by comparing the pre- and post-training MCQ test performance of each individual participant (irrespective of training modality).

2) *The effectiveness of ViTAL training*. This was measured by comparing the performance of the two cohorts in the post-training MCQ test.

The Shapiro-Wilk test showed that the pre- and post-training test data were not normally distributed, therefore non-parametric tests were applied with data presented as medians and interquartile ranges (IQR). The MCQ test results are expressed as a percentage score. The Wilcoxon signed rank test was applied to make paired comparisons between the pre- and post-training test scores for each individual participant in the Standard and ViTAL Training cohorts. The test performance of the Standard Training and ViTAL Training cohorts at the pre-training and post-training stage were compared using the Mann-Whitney *U* test.

All data were analysed with SPSS software. A p-value < 0.05 was considered statistically significant.

4.3. RESULTS

4.3.1. Difficulty in teaching the surgical steps of the OUKA procedure

Fifty-six experts were contacted via email and 34 (61%) responded to the survey (see Chapter 2 for details of the Delphi process methodology). Table 4-1 shows the opinion of OUKA experts with regard to which aspects of the procedure are easy to teach orthopaedic trainees. The responses were carefully considered during the development of ViTAL's contents and efforts were made to further clarify and emphasise the surgical steps that the survey respondents found difficult to teach.

Key surgical steps of the Oxford unicompartmental knee arthroplasty procedure as identified by the Delphi survey	Is this step easy to teach orthopaedic trainees?	
	Yes	No
Patient positioning	31 (94)	2 (6)
Skin incision and surgical approach	32 (97)	1 (3)
Joint inspection and assessment for contra-indications	29 (88)	4 (12)
Osteophyte excision	32 (97)	1 (3)
Seating of tibial saw guide onto tibia and estimation of resection height	13 (39)	20 (61)
Execution of the vertical cut for tibia	20 (61)	13 (39)
Execution of the horizontal tibia cut and removal of tibial plateau	19 (58)	14 (42)
Ensuring sufficient amount of tibial plateau is excised and further resection made if necessary	24 (73)	9 (27)
Assembly of the femoral drill guide and placement of the femoral drill holes	17 (52)	16 (48)
Excision of the posterior femoral condyle with saw using saw guide	25(76)	8 (24)
Milling and measurement of flexion / extension gaps	25 (76)	8 (24)
Estimation of the trial tibial template	31 (94)	2 (6)
Preparation of the tibial plateau keel slots	27 (82)	6 (18)
Selection of meniscal bearing and ligament balancing	23 (70)	10 (30)
Cementation of the implant	19 (58)	13 (42)
Attention to MCL / soft tissues	27 (82)	6 (18)
Final inspection for bearing impingement and alignment	27 (82)	6 (18)

Table 4-1: Responses of the survey participants with regards to how difficult it is to teach the key steps of the Oxford unicompartmental arthroplasty procedure (n=33). Percentages are given in parentheses.

4.3.2. Usability study

Table 4-2 summarises the demographics of participants who evaluated the usability of ViTAL. Although 21 of the participants stated that they played computer games, only two of these played games on a daily basis. Twelve participants stated that they never played computer games. None of the medical students or biomedical scientists had any previous experience of observing or assisting in UKA surgery, however, experience of performing or assisting in the OUKA procedure was recorded for clinicians and theatre staff. Seven participants had performed at least one OUKA procedure as the primary surgeon. Eighteen (54.5%) of the participants reported assisting in more than one OUKA procedure. Six (18.2%) had assisted in more than 100 procedures, three (9.1%) had assisted in 31-100 procedures, three (9.1%) had assisted in 10-30 procedures and six (18.2%) had assisted in less than 10 procedures.

Usability Survey Demographics	Participants (n = 33)
Mean Age (Years) (Range)	30.3 (22 – 52)
Male : Female	24 : 9
Occupation: • Expert UKA surgeons • Orthopaedic Higher Surgical Trainees • Foundation Doctors • Operating room scrub staff • Fifth-year medical students • Biomedical scientists	 4 6 6 3 10 4
Daily work involves regular computer use: • Yes • No	 30 3
Level of computing knowledge: • Never use computers • Email / internet / social networking / entertainment • Word processing / presentations • Database entry / manipulation • Scientific use (statistical analysis / equation solving) • Graphics design / video editing / music composition • Software development / programming	 0 2 6 7 14 1 3
Frequency of computer game use: • Never • Monthly • Weekly • Daily	 12 14 5 2
Genre of games played	See Appendix 5
Previous use of VR simulators or serious games: • Yes • No	 2 31
Previous experience of observing or assisting in UKA surgery	See text

Table 4-2: Demographics of participants evaluating the usability of the ViTAL cognitive simulator.

Table 4-3 summarises the results of the software usability questions completed by all 33 participants. In addition, 67% (12/18) of the clinicians and operating room staff (who were asked about the representation of the OUKA procedure by ViTAL), agreed/strongly agreed with the statement that the *"virtual representation is a realistic view of the procedure"*. Furthermore, 89% (16/18) of these individuals agreed/strongly agreed that the *"tasks reflect those encountered in real-life surgery"*. All of the clinicians and operating room staff agreed/strongly

agreed that they would *"use this software as a supplementary tool for learning/teaching the OUKA procedure"*, and that the *"software teaches the procedural steps you need to perform the OUKA"*.

Software Usability Questions	Likert scale Rating Median (Range)
Software interface was too complex	2 (1 to 5)
The software is slow / unresponsive to your actions	2 (1 to 4)
Textual information was not sufficient	2 (1 to 3)
It was difficult to associate the textual description of the procedure with the graphics	2 (1 to 4)
Additional "voice-over" information would make the instructions easier to follow	4 (2 to 5)
The procedure should be displayed / viewed in stereoscopic 3D	3 (1 to 5)
This software should be accessible online for remote use 24 hours a day	4 (3 to 5)
I found it enjoyable to learn using the software	4 (3 to 5)

Table 4-3: Opinions of participants with regards to the usability of the ViTAL cognitive simulator. Rating based on 5-point Likert scale (1 = strongly disagree to 5 = strongly disagree).

4.3.3. Determining the effectiveness of ViTAL

Demographics data (Table 4-4) demonstrated that there were no differences between the two cohorts based on age, sex, computer game use, and future career choice. All participants used computers as part of their daily work and none of them had previously observed or assisted in UKA surgery. One participant in the ViTAL cohort (who was previously a trader in the London Stock Exchange) had experience of using VR-based simulation software and "serious games" for educational purposes.

Demographics	Standard Cohort (n = 12)	ViTAL Cohort (n = 13)
Mean Age (Years) (Range)	22.9 (22-27)	23.0 (22-28)
Male : Female	5 : 7	7 : 6
Daily work involved regular computer use • Yes • No	 12 0	 13 0
Frequency of computer game use : • Never • Monthly • Weekly • Daily	 4 7 1 0	 3 7 2 1
Previous use of VR simulators or serious games: • Yes • No	 0 12	 1 12
Future career choice: • Surgery • Not surgery • Undecided	 4 7 1	 4 6 3
Previous experience of observing or assisting in UKA surgery	Nil	Nil

Table 4-4: Demographics of the medical students taking part in the randomised controlled trail to assess the effectiveness of the ViTAL cognitive simulator. Results are not statistically significant.

The effect of the two training modalities is represented in Figure 4-5. The pre-training test showed no differences in the baseline cognitive skills of the two groups (Mann-Whitney *U* test, p=0.166). The median test score in the Standard Training cohort was 30% (interquartile range, or IQR, 20 to 30) compared with 20% (IQR 10 to 30) in the ViTAL Training cohort. Pairwise comparisons demonstrated that both training modalities result in statistically significant improvement in cognitive skills (Wilcoxon Signed Rank test, $p<0.0001$). However, the ViTAL Training cohort performed significantly better in the post-training MCQ test in comparison to the Standard Training cohort (Mann-Whitney *U* test, $p<0.0001$) demonstrating better cognitive skills acquisition. The median

test score in the Standard Training cohort was 53% (IQR 46 to 64) compared with 85% (IQR 80 to 93) in the ViTAL Training cohort.

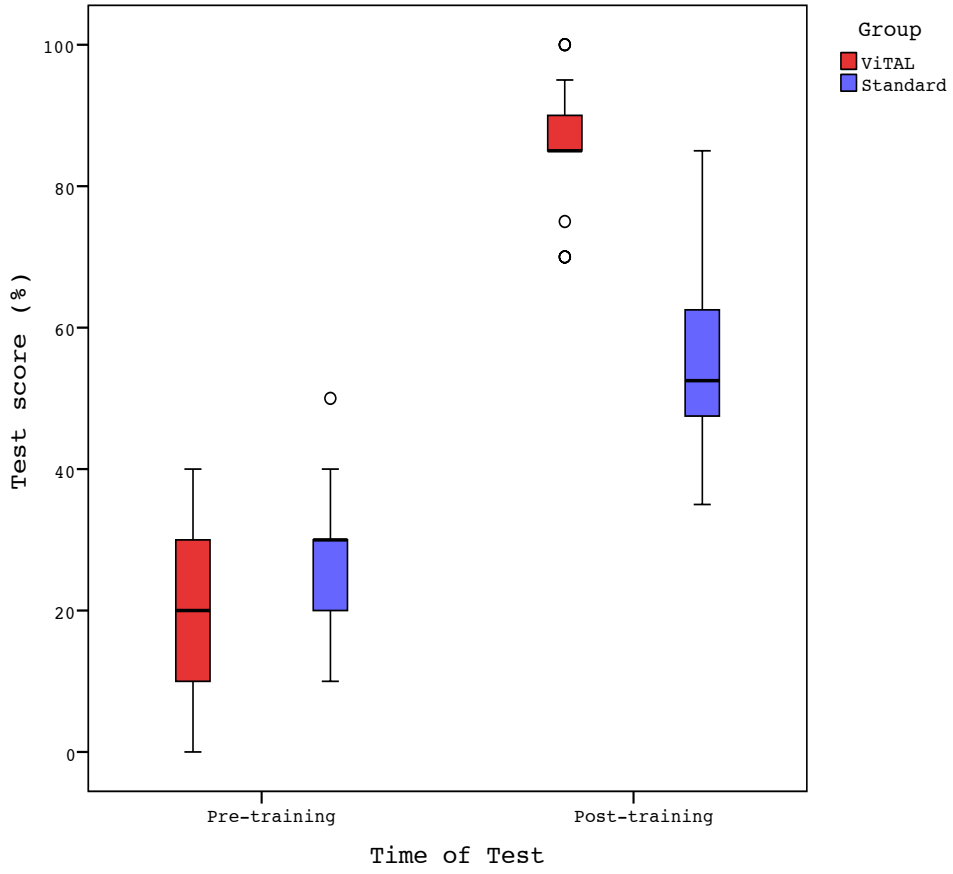

Figure 4-5: Performance of the medical students in the pre- and post-training MCQ tests for assessment of the cognitive skills necessary to perform the OUKA procedure (box represents the interquartile range, bold line in the box represents the median, whiskers represent the 2.5 and 97.5 percentiles, and circles are outliers).

4.4. DISCUSSION

4.4.1. Implications of the study

This chapter outlines the successful development and usability testing of a VR-based simulator for teaching the cognitive skills necessary to perform the OUKA procedure. More importantly, the training effectiveness of ViTAL was assessed in a randomised controlled trial demonstrating that, if used alongside existing teaching modalities, it can result in better acquisition of cognitive skills in novice trainees (medical students).

The influence of cognitive skills on surgical ability has long been emphasised, with Spencer famously stating that " a skilfully performed operation is about 75% decision making and 25% dexterity"[293]. In another commonly quoted study, Cuschieri et al.[294] surveyed the opinion of "master" surgeons with regards to important attributes of surgical trainees and found that 70% of the respondents rated "cognitive ability" amongst the top three attributes (with "innate dexterity" and "personality" as the remaining two). In spite of this, the focus of surgical education over the past two decades has been heavily weighted towards the development of VR simulators that are designed to improve technical skills, and has been shown to be successful in orthopaedic surgery[152] and other surgical disciplines[295]. Although it is essential to develop simulators for technical skills training, the key role of cognitive skills (in addition to other non-technical skills) must not be overlooked. This is supported by a growing body of evidence that emphasises the importance of non-technical skills in surgery[283,296-298].

Even though there is encouraging evidence to justify a call for improving the cognitive skills of surgeons outside the real operating room environment, the cost

of purchasing VR simulators and, more importantly, the level of access to them by trainees remains a barrier to their wide-scale use. It is, therefore, important to develop cognitive simulators that are cost effective and can be accessed easily by learners. Subramanian et al.[299] were able to demonstrate that a low fidelity web-based cognitive simulator was significantly more effective at teaching surgical trainees about cardiac conditions compared to traditional PowerPoint lectures (Microsoft Corporation, Redmond, Washington). They demonstrated that a highly focused, but interactive simulator, could be used to engage the learners in patient management scenarios that resulted in better acquisition of clinical knowledge. In developing the ViTAL software, the aim was to ensure that it could be utilised by as many learners as possible. Consequently, it was designed to have cross-platform compatibility (Windows / Mac / Linux) and can be installed from a CD-ROM for use on a standard desktop computer without the need for expensive hardware such as haptic (tactile feedback) technology. This *extensibility* means that its' use is not limited to specialised surgical skills laboratories of large institutions, and can potentially dramatically increase ViTAL's educational impact. Furthermore, it is important to provide modern-day surgical trainees with software-based educational tools that are accessible 24 hours a day for independent study, because they face a growing number of novel and challenging surgical procedures that require mastery during short training programmes. The need for such educational tools is further reflected by the recent success of mobile apps such as Touch Surgery that aim to teach common surgical procedures by breaking them down into their component steps[300].

Feedback from the participants during the usability study was used to enhance several aspects of ViTAL. Major changes involved the addition of a voice-over

and removal of the 3D/stereoscopic viewing option. The usability study demonstrated that across all surveyed user groups, the majority of participants found ViTAL enjoyable and easy to use. The expert group was overwhelmingly positive in their evaluation of ViTAL's realism and relevance as a training tool, providing a strong indication of the positive future potential of ViTAL. This is important because an educational tool is of limited use if the learner does not accept or enjoy using it.

The content of any simulator has a major impact on its ultimate success as a training tool. Furthermore, there are often key steps in surgical procedures that have a major impact on successful clinical outcome, in addition to specific steps, that are particularly difficult to teach novices. The key steps of the OUKA procedure were identified using the Delphi technique (as described in Chapter 2). As a result aspects of the procedure that were thought to be more difficult to teach trainees were emphasised in ViTAL. However, the pre- and post-training tests were designed specifically to assess information that was common to both training modalities, hence there was no bias in favour of either training modality. This suggests that ViTAL was able to convey the same information that was present in the surgical technique manual and DVD in a more effective manner. A plausible explanation for this may be that learner inactivity occurred when using the surgical technique manual or DVD, whereas the higher level of interaction employed by ViTAL meant that the users were more receptive to learning.

The key aim of ViTAL was to improve the cognitive skills necessary for OUKA – in particular the procedural steps of this operation. This was considered an important area for improvement because previous studies in other surgical domains have also demonstrated that surgical trainees have particular difficulties

in this domain. Tang et al.[301] discovered that approximately 50% of errors enacted by surgical trainees during a skills training course for laparoscopic cholecystectomy were related to inadequate knowledge related to procedural steps of the surgery. These involved either omission of important steps or execution of steps in the wrong sequence. Thus, VR simulators that complement or improve existing methods of teaching these cognitive skills are in great demand at a time when external factors are having a negative impact on the delivery of surgical education. Another study, that investigated the impact of attendance at an OUKA course on the improvement of cognitive skills, highlighted the fact that some participants were unable to answer a number of key questions relating to cognitive components of the procedure even after attending the course[302]. These major knowledge gaps could potentially affect the future performance of the course attendees during real-life OUKA procedures. Such studies provide further evidence that current teaching modalities must be further improved and, therefore, VR-based cognitive simulators seem to have the potential to help bridge this gap.

It would seem logical that by improving the cognitive skills necessary for surgical procedures outside of the real operating room, more time can be dedicated to improving other domains such as technical skills inside the operating room, thereby increasing the efficiency of surgical training. However, although the findings of this randomised study are extremely encouraging, they do not automatically infer that cognitive skills training with ViTAL results in better technical performance during OUKA surgery. The next logical step is to determine the effectiveness of ViTAL on the technical skills of orthopaedic trainees.

4.4.2. Limitations

This study had some limitations. It was not possible to perform a long-term retention test to determine how much of the knowledge acquired by the two types of intervention was retained in the future. This was logistically difficult to execute since the accuracy of this retention test may have been affected by cross-contamination of the groups due to exposure of the students to additional OUKA procedures during the course of their orthopaedic attachment. Indeed, a deliberate attempt was made prevent this by conducting the post-training tests within 24 hours of the third (final) training session.

Some may argue that the validity of the MCQs could have been improved by utilising techniques such as the Angoff method[303] of standard setting, whereby cut-scores for minimum competency are obtained. These methods would not have been appropriate in the current study as the aim was to compare the relative degree of cognitive skills improvement by the participants in each cohort, as opposed to whether or not a specific threshold was reached.

Another limitation is that allowing the ViTAL cohort access to the surgical technique manual and DVD may have provided them with a greater advantage, and resulted in the better Post-training performance. A deliberate decision was taken to allow the ViTAL cohort access to this material, as this would be a fair reflection of a 'real-life' scenario for surgical trainees. Trainees regularly use textbooks and DVDs to learn about surgical procedures and denying access to these sources would have created an artificial learning environment. Thus, it is currently reasonable to view ViTAL as an educational tool for augmenting traditional learning modalities.

4.4.3. Summary

This chapter has sought to apply a number of different methods to address an important challenge that currently faces surgical educators – the improvement of cognitive skills outside the real operating theatre. The surgical procedure selected to demonstrate this proof of concept was the commonly performed but technically challenging OUKA, whilst the method of improving cognitive skills was VR-based surgical instruction. The opinion of worldwide experts was sought in order to determine the content of the simulator. The simulator platform was designed and implemented through collaboration with computer science and graphic design experts. Undertaking a usability study across a wide range of end-users ensured optimisation of the face validity and usability of the simulator. Finally, a randomised controlled trial clearly demonstrated the effectiveness of the ViTAL simulator in improving the cognitive skills necessary for performing OUKA. This development model could be employed to design and develop cognitive simulators for improving performance during other challenging surgical procedures. Furthermore, the results enable the planning of further studies to determine the effect of cognitive skills training on technical performance during OUKA.

5. RANDOMISED CONTROLLED TRIAL TO DETERMINE THE IMPACT OF COGNITIVE SKILLS TRAINING USING THE ViTAL SIMULATOR ON PERFORMANCE DURING SIMULATED UNICOMPARTMENTAL KNEE ARTHROPLASTY

5.1. INTRODUCTION

There is no doubt that one of the most important factors that impacts on the outcome of a surgical procedure is the technical skill of the operating surgeon. For this reason, the major focus of simulation-based surgical technology has been directed towards teaching technical skills. This has resulted in a comprehensive body of evidence that supports the value of practicing psychomotor skills on surgical simulators[304-306]. However, the ability to perform the technical manoeuvres relating to a surgical procedure cannot guarantee optimum clinical outcome. The surgeon's overall performance is also strongly influenced by cognitive factors such as decision-making, procedural knowledge, error detection and surgical planning. The effect of cognitive skills training using simulators has received comparatively little attention. Some promising work has taken place in relation to this within orthopaedic surgery by Sabri and colleagues[307], who developed a cognitive simulator in the form of a "serious game" to teach TKA. Although usability testing of this cognitive simulator has been successful, no work has yet been published to demonstrate its value in terms of training effectiveness or transfer validity. Outside orthopaedics, Loveday et al.[286] were able to show that a cognitive simulator for teaching laparoscopic appendectomy could improve the cognitive skills of junior surgical trainees, but the study did not progress to

investigate the effect of the cognitive simulator on surgical performance. Van Herzeele et al.[308] provided further evidence for the effectiveness of cognitive skills training by demonstrating that didactic cognitive skills tuition by an expert on an endovascular VR simulator can significantly improve the quality of the end product on the simulator. One drawback of the methods of training used in this study was that it required the constant presence of an expert tutor to provide instruction. The educational value of VR simulators that teach cognitive skills without the need for the presence of an expert tutor will be far greater – especially if this training ultimately results in an improvement in surgical performance.

The previous chapter described the stepwise development of ViTAL, and successfully demonstrated that ViTAL training was more effective than standard training modalities for improvement of the cognitive skills necessary for the OUKA procedure. However, improvement of this so-called declarative knowledge, i.e. "knowing what to do", in a population of novices may not automatically translate to an improvement in actual surgical performance, i.e. "showing how to do it". The next logical step was therefore to determine whether or not the ViTAL-based cognitive skills improvement had any effect on the technical performance of the target population for which the simulator was designed, i.e. orthopaedic trainees.

5.1.1. Aims:

The purpose of this chapter was, therefore, to determine the effect of cognitive skills training using the ViTAL simulator on the ability of orthopaedic trainees to perform the OUKA procedure on anatomic dry-model bones in a surgical skills laboratory environment.

5.2. METHODS

5.2.1. Setting and participants

This randomised controlled trial took place in the Oxford Orthopaedic Simulation and Education Centre over a two-month period. Institutional approval was obtained and all participants recruited to the trial gave informed consent. The participants consisted of first, second and third year orthopaedic trainees. In order to be eligible for recruitment, study participants had to demonstrate surgical logbook evidence of having performed at least five independent TKA procedures (under supervision of a consultant). This inclusion criterion ensured that they possessed the basic psychomotor skills for performing total joint arthroplasty. Participants who had previously performed UKAs, or were working for an orthopaedic firm that regularly undertook UKAs at the time of the trial, were excluded. These exclusion criteria were imposed in order to maintain the homogeneity of the participants and to prevent 'unfair' exposure of some participants to additional OUKA educational opportunities during the study period.

Sixteen orthopaedic trainees were recruited for the study. They completed a questionnaire to determine basic demographic data, previous surgical experience, perceived level of computing knowledge, frequency of computer game use, and previous exposure to 'serious games'. Participants were subsequently randomised, using sealed opaque and consecutively numbered envelopes, to undergo either Standard Training (control, n = 8) or ViTAL Training (intervention, n = 8). They were asked not to disclose their randomisation status to anyone other than the primary investigator. Figure 5-1 illustrates the flow of participants through the

trial.

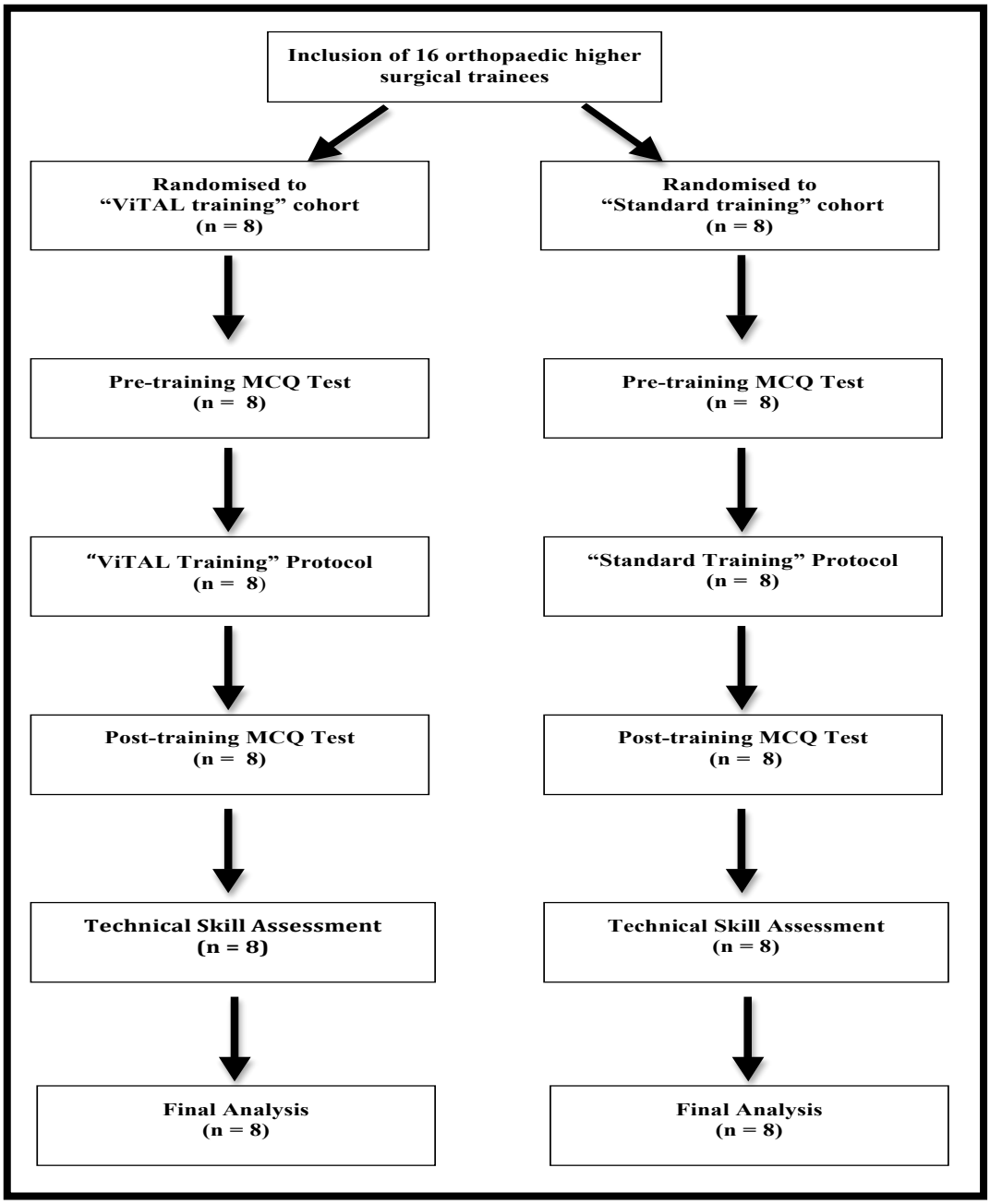

Figure 5-1: Flow diagram of the participants (orthopaedic trainees) through the trial.

5.2.1.1. Cognitive skills training

This was the same training protocol as used in the previous study, which had investigated the effects of ViTAL and Standard training in a medical student population (Chapter 4). Both cohorts used the OUKA surgical technique manual

and operative procedure DVD (Biomet Inc., Warsaw, Indiana, USA). The ViTAL Training cohort used the ViTAL simulator *in addition* to the latter two learning sources. As previously described, orthopaedic trainees who first embark on learning about the OUKA procedure usually begin by using the surgical technique manual and DVD as a source of information for private study. The previous study involving medical students had demonstrated that both teaching modalities resulted in improved learning. However, the degree of improvement in the ViTAL Training cohort was greater than the Standard Training cohort. Participants in the current study attended the department's simulation centre once a week (for a total of three weeks) for one of two methods of cognitive skills training. On their first training session, the ViTAL Training cohort received a standardised 5-minute briefing for the simulator, outlining its proposed objectives and user instructions. Both cohorts were required to read the surgical technique manual in their own time (at least once per week). In order to address the ethical dilemma of the participants in the Standard Training cohort being denied access to a potentially beneficial educational opportunity, these individuals were offered the chance to train on the ViTAL simulator at the end of the study period. Both cohorts continued with their normal orthopaedic registrar training during the study period, but were asked to document any additional exposure to OUKA surgery.

5.2.1.2. Assessment of cognitive skills

Cognitive skills relevant to the OUKA procedure were assessed using pre- and post-training MCQ tests that had been validated during the development of the ViTAL cognitive simulator (Chapter 4). The tests consisted of twenty single best answer MCQs with four options for each stem (Appendix 6). The results were expressed as percentage scores. The questions were designed to test decision-

making skills and procedural knowledge necessary for carrying out OUKA surgery. The topics that were examined were covered by both training protocols. The pre-training MCQ test took place prior to the first training session. The post-training MCQ test took place within one day of the third (final) training session. The questions were marked by one of the investigators who was blinded to the identity and training allocation of the participants. The post-training tests were performed within one day of completing the final training session.

5.2.1.3. Post-training assessment of technical skills

The aim of assessing technical skills was to evaluate whether or not cognitive skills training would have an effect on technical performance during a simulated OUKA procedure. This was conducted by evaluating the ability of the participants to perform the procedure on a benchtop knee simulator in a surgical skills laboratory environment. All assessments were performed within one day of completing the final training session.

Equipment set-up and procedure:

A benchtop knee simulator (model 1150-2, Sawbones Europe, Malmo, Sweden) with its accompanying soft tissue cover (model 1506-4 and 1506-5) was set up in a surgical skills laboratory (Figure 5-2). The models simulated a knee with varus pattern of osteoarthritis with its associated pathological changes such as medial compartment wear and osteophytes. The simulator was mounted on a custom-made rig to represent the special leg holder for the real OUKA procedure. The instrumentation consisted of the standard surgical equipment used in OUKA (Biomet, Swindon, United Kingdom). Participants were required to carry out the OUKA procedure in its entirety with a descriptive narrative of their actions. This

was to ensure that the assessors could judge their thought processes and decision-making skills during the procedure. They were instructed to treat the simulated scenario as a 'real-life case' that was being assessed by a consultant. The only surgical steps that could not be performed on the benchtop simulator were 'pre-operative planning and contra-indications', 'patient positioning', and 'surgical exposure of the joint' since they had already been pre-determined. Participants were, therefore, required to provide a detailed description of these steps at the start of the procedure. The primary investigator acted as the first assistant but did not provide any specific guidance.

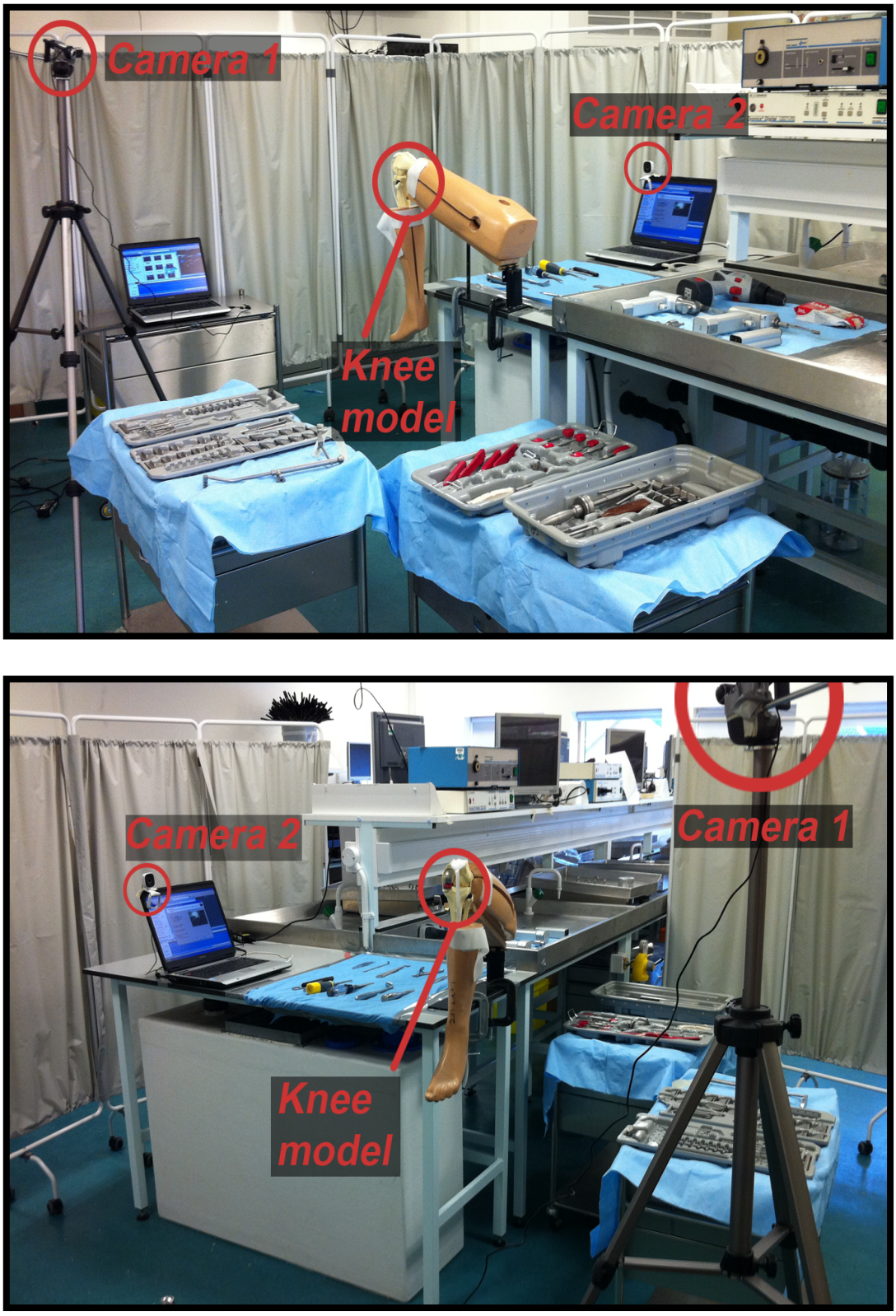

Figure 5-2: Set-up of OUKA instrumentation, knee simulator, and webcams for evaluation of technical skills after training.

Data acquisition and assessment of technical skills:

All procedures were digitally recorded using two high-resolution webcams with sound recording capability. Two differing views of the operating field were deliberately recorded in order to maximise data capture in the event of the surgeon inadvertently obscuring one of the cameras during a particular surgical step (Figure 5-2). The recordings were anonymously coded and stored as QuickTime video files (version 10.0, Apple Inc, Cupertino, California, USA) in order to be scored by an independent assessor. Performance was evaluated using the Oxford Arthroplasty Technical Skill Score (OATSS) system (Figure 5-3). OATSS has been specifically developed for the assessment of technical skills during OUKA surgery, and has been shown to possess construct validity (Chapter 2, Study C). The 16 pairs of videos were viewed in their entirety by a blinded assessor who scored the performances. The assessor was a consultant knee surgeon with fellowship training in OUKA surgery and also previous experience of using the OATSS scoring system. Participants were instructed to conceal their randomisation status from the assessor. The primary investigator also viewed and scored all the procedures using OATSS in an attempt to gain further information about the inter-observer reliability of using this scoring system within the surgical skills laboratory environment. However, the results of the primary investigator's scores were only used to determine inter-observer reliability, and not used to make direct comparisons between the two study cohorts. This was because the primary investigator was not blinded to the participants' training allocation. Nevertheless, it was felt that this was a suitable method of estimating inter-observer reliability, given the feasibility constraints of identifying suitably trained individuals who

could partake in the lengthy assessment process (viewing over 30 hours of video footage).

The Oxford Arthroplasty Technical Skill Score (OATSS) evaluation form

Surgical Step	1	2	3	4	5
Patient positioning	Unable to perform / Unsafe		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Skin incision and surgical approach	Unable to perform / Incorrectly placed incision / Unsafe approach		Performed with difficulty but no safety concerns		Accurate, safe & skillful approach
Joint inspection & assessment for contra-indications	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurate & skillful assessment
Osteophyte excision	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Seating of tibial saw guide onto tibia and estimation of resection height	Unable to perform / Over excision which compromised procedure		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Execution of the vertical cut for tibia	Unable to perform / Incorrect med-lat position / Unsafe (e.g. saw blade handle raised / plunged deeply risking posterior cortex damage)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Execution of the horizontal tibia cut and removal of tibial plateau	Unable to perform / Unsafe (e.g. MCL not protected)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Ensuring sufficient amount of tibial plateau is excised and further resection made if necessary	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed. Did not require "re-cut"
Insertion of the femoral IM rod with attention to patella	Unable to perform / Neglected to perform / Unsafe (e.g. caused damage to patella)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Assembly of the femoral drill guide & placement of the femoral drill holes	Unable to perform / Neglected to perform all alignment checks / Incorrectly performed		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Excision of the posterior femoral condyle with saw using saw guide	Unable to perform / Neglected to perform / Unsafe (e.g. saw blade plunged too deeply)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Milling and measurement of flexion / extension gaps	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Estimation of the trial tibial template	Unable to perform / Neglected to perform / Incorrect sizing with large medial overhang of template		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Preparation of the tibial plateau keel slots	Unable to perform / Neglected to perform / Unsafe clearing of keel slot		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Selection of meniscal bearing & ligament balancing	Unable to perform / Neglected to perform / Unsafe (e.g. risked stretching / avulsion of ligaments)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed
Cementation and/or implantation technique	Unable to perform / Unsafe (e.g. allowed cement to harden befor implantation was complete)		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed. Limb in correct position during implantation
Attention to MCL / soft tissues	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		All carefully protected at all times.
Final inspection for bearing impingement and alignment	Unable to perform / Neglected to perform		Performed with difficulty but no safety concerns		Accurately, safely & skillfully performed

Figure 5-3: The Oxford Arthroplasty Technical Skill Score evaluation form used by the expert evaluators to assess the performance of participants during the simulated OUKA procedure. Scores for each of the 19 items were summed to generate the total OATSS value (maximum score: 90).

5.2.2. Outcomes and statistical analysis

Target sample size:

As the primary aim of this study was to determine the effect of cognitive skills training on orthopaedic trainees' technical ability to perform OUKA, it was not appropriate to use the results of the medical student study (Chapter 4) for the purpose of sample size calculation. However, previous validation studies during the development of OATSS had demonstrated that the mean OATSS score for orthopaedic trainees performing OUKAs *in vivo* was 63.15 points, with a standard deviation of 8.69 points (Section 2.4.2). Furthermore, previous studies in the field of surgical skills research, which had used global rating scales and checklists to compare simulator training with control groups, had reported effect sizes ranging between 0.8 to 1.92 SDs[280,309,310]. An improvement/effect size of 1.5 standard deviations (equating to 13 points on the OATSS scoring system) was therefore deemed appropriate. Using an alpha of 0.05, and a power of 0.8, it was necessary to recruit at least 8 participants in each cohort.

The primary outcome measure was:

- To determine whether ViTAL Training is more effective than Standard Training in improving technical skills within a skills laboratory setting – assessed by comparing OATSS scores of the two cohorts.

The secondary outcome measures were:

- To determine whether training can improve cognitive skills (irrespective of the *type* of training) – assessed by comparing the pre- and post-training

MCQ test performance of each individual participant.

- To determine whether ViTAL Training is more effective than Standard Training in improving cognitive skills – assessed by comparing the performance of the two cohorts in the post-training MCQ test.

The Shapiro-Wilk test showed that the pre- and post-training test data were not normally distributed, therefore non-parametric tests were applied with data presented as medians and interquartile ranges (IQR). The Wilcoxon signed-rank test was applied to make paired comparisons between the pre- and post-training test scores for each individual participant within the Standard and ViTAL Training cohorts. Comparisons between the two cohorts were made at the pre-training and post-training stages using the Mann-Whitney *U* test. Parametric analysis using an independent group *t*-test was used to compare the technical performance of the cohorts based on their OATSS scores. The inter-rater reliability of the overall OATSS score was assessed using the intraclass correlation coefficient (ICC), with ICC > 0.61 being considered substantial agreement. Inter-rater reliability for sub-sections of OATSS was also assessed using the Weighted kappa coefficient[237], with $k > 0.6$ being considered substantial agreement. Cohort demographics were compared using either the independent group *t*-test or Fisher's Exact test as appropriate. All data were analysed with SPSS software. A p-value < 0.05 was considered statistically significant.

5.3. RESULTS

5.3.1. Cohort demograpchics

The two cohorts were evenly matched in terms of general demographics, frequency of computer-game use, and surgical experience (Table 5-1). None of the participants had attended an instructional course for OUKA surgery, but they had all assisted in real-life OUKA procedures. One of the participants in the ViTAL Training cohort had previous experience of VR simulator training - though this training was for laparoscopic surgery.

Demographics	Standard Cohort (n = 8)	ViTAL Cohort (n = 8)
Mean Age (Years) (Range)	32 (29-34)	31.4 (29-33)
Male : Female	8 : 0	8 : 0
Daily work involves regular computer use: • Yes • No	 8 0	 8 0
Frequency of computer game use: • Never • Monthly • Weekly • Daily	 3 2 2 1	 2 3 1 2
Previous use of VR simulators or serious games: • Yes • No	 0 8	 1 7
Mean TKA procedures preformed as primary surgeons (Range)	14.6 (7-29)	13.1 (6-20)
Mean OUKA procedures performed as primary surgeon	0	0
Mean OUKAs procedures assisted (Range)	19.25 (8-41)	18.63 (7-49)
Attendance at the OUKA instructional course: • Yes • No	 0 8	 0 8

Table 5-1: Comparison of demographics between orthopaedic trainees in the Standard Training and ViTAL Training cohorts. All values were not significantly different between groups ($p > 0.05$).

5.3.2. Determining the effectiveness of ViTAL simluator training for improving cognitive skills

There was no significant difference in the baseline cognitive skills of the two cohorts, as evident from the pre-training MCQ test (Mann-Whitney *U* test, p=0.549). The median test score in the Standard Training cohort was 70% (interquartile range (IQR) 66.25 to 80) compared with 75% (IQR 70 to 87.5) in the ViTAL Training cohort (Figure 5-4). Pairwise comparisons demonstrated that both training modalities resulted in statistically significant improvements in cognitive skills (Wilcoxon signed-rank test, p<0.001). However, the ViTAL Training cohort demonstrated better cognitive skills acquisition, as evident by their significantly higher post-training MCQ test results (Mann-Whitney *U* test, p<0.05). The median test score in the Standard Training cohort was 87.5% (IQR 85 to 93.75) compared with 97.5% (IQR 95 to 100) in the ViTAL Training cohort.

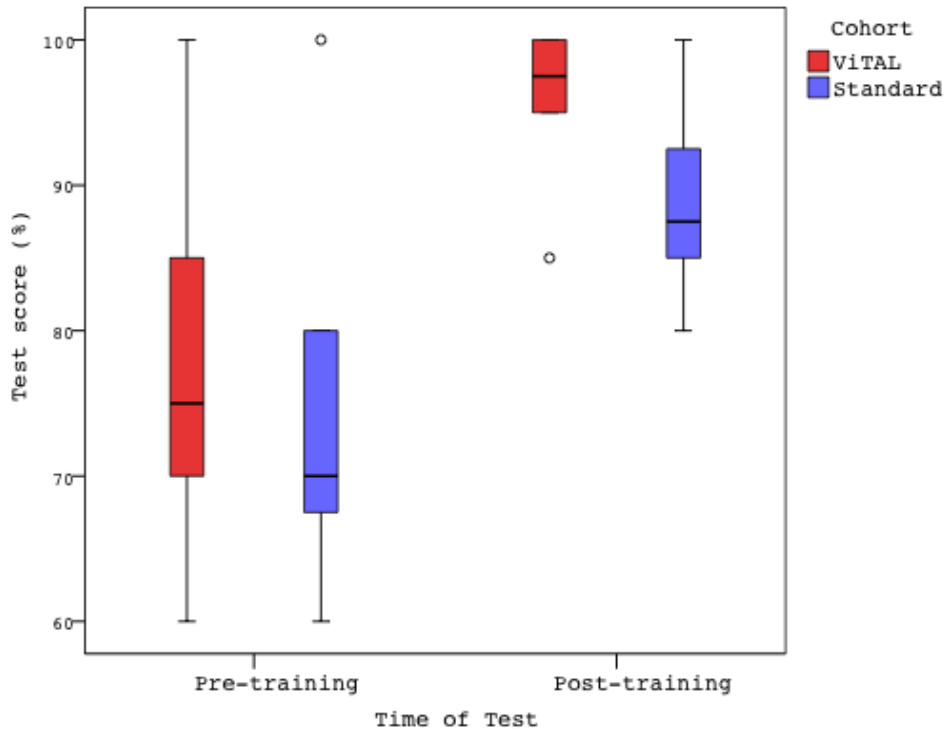

Figure 5-4: Performance of the orthopaedic trainees in the pre- and post-training MCQ tests for assessment of the cognitive skills necessary to perform the OUKA procedure (box represents the interquartile range, bold line in the box represents the median, whiskers represent the 2.5 and 97.5 percentiles, and the circles are outliers).

5.3.3. Determining the effectiveness of ViTAL simluator training for improving technical skills

Figure 5-5 illustrates the results for post-training assessment of OUKA performance in the surgical skills laboratory setting. The ability of the ViTAL Training cohort to perform OUKA in the surgical skills laboratory was significantly better than the Standard Training cohort ($p = 0.034$). The mean OATSS score of the ViTAL Training cohort was 79.6 (SD 6.5) compared with a mean OATSS score of 71.8 (SD 6.8) for the Standard Training cohort. The improvement gained from cognitive skills training using the ViTAL simulator amounted to an effect size of 1.18 (Cohen's *d*). The inter-rater reliability for the OATSS in the simulation laboratory setting was excellent (ICC = 0.96 with 95% confidence intervals 0.90-0.99). The inter-rater reliability for the sub-sections was also satisfactory (weighted kappa coefficient, $k = 0.62$).

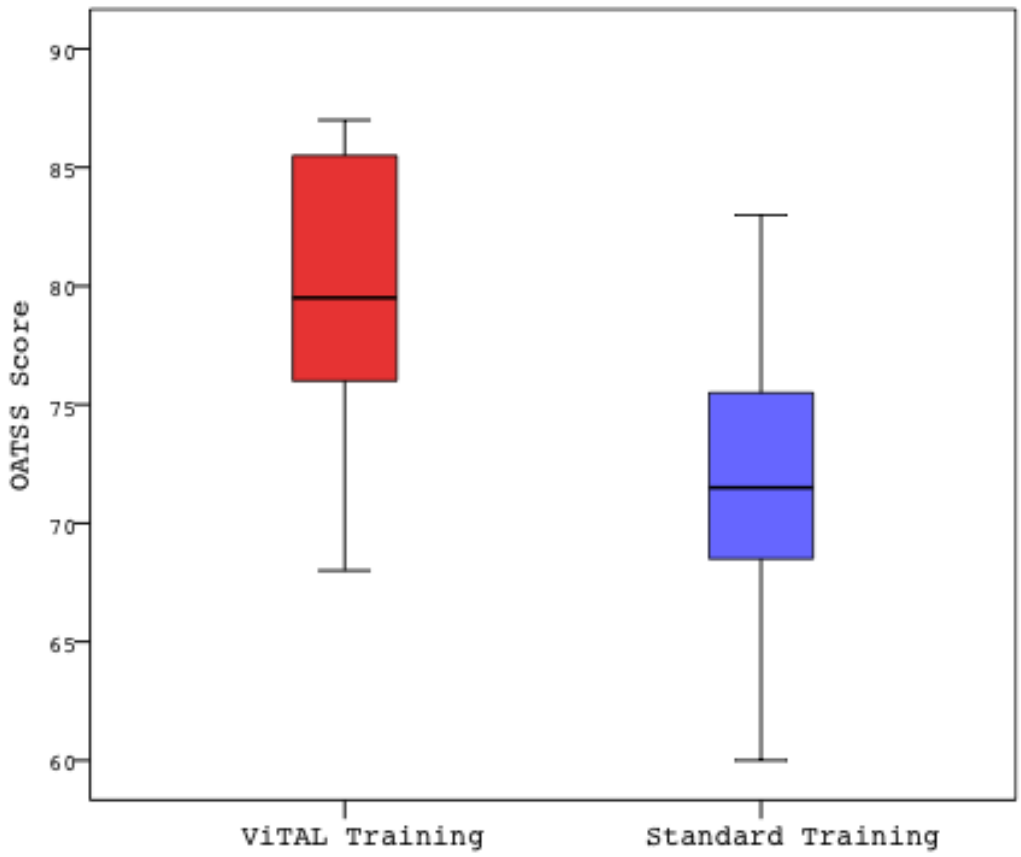

Figure 5-5: Post-training assessment of OUKA performance using the Oxford Arthroplasty Technical Skill Scale (OATSS) scoring system (maximum possible score = 90, box represents the interquartile range, bold line in the box represents the median, and the whiskers represent the 2.5 and 97.5 percentiles)

5.4. DISCUSSION

5.4.1. Implications of the study

This randomised controlled study demonstrated that cognitive skills training of orthopaedic trainees using the ViTAL simulator not only results in the better acquisition of cognitive skills, but also results in better technical skill performance during OUKA procedures on anatomic dry-model bones. The most important aspect of this finding is that the simulator-based training focused on improving cognitive skills, and no technical skills training was provided. To our knowledge, this is the first study in the orthopaedic literature that has been able to demonstrate

the effectiveness of a purely cognitive VR-based simulator for improving technical skills during arthroplasty surgery. In addition, the findings provide further evidence of the importance of cognitive skills for overall surgical performance during OUKA surgery.

Research from a wide range of surgical disciplines has highlighted the strong influence of cognitive skills on surgical performance. Meyerson et al.[311] recently investigated the type of errors committed by trainees during simulated thorascopic lobectomy – a technically demanding procedure with a steep learning curve. They discovered that, although the potential number of cognitive and technical errors that could have been made during the procedure was similar, the proportion of cognitive errors made by trainees was significantly higher than technical errors. Simulation-based studies in laparoscopic surgery have also highlighted that a large proportion of intra-operative errors are due to inadequacies in cognitive skills, such as the omission of key surgical steps[301] or poor decision-making[312]. The authors of the latter study even suggested that "...training and feedback on motor skills and manual dexterity would not have helped these residents complete the hernia repair". These observations have been previously investigated in depth[55] and can largely be explained by the Cognitive Load theory[313]. This is based on the fact that the human mind has a limited amount of processing capacity, combined with partly independent processing units available for visual and auditory information. When these systems are overwhelmed with information, the individual will eventually be rendered unable to multitask. In the case of surgery, this may manifest itself as the inability to remember a sequence of surgical steps or make a particular decision about a critical set of intra-operative actions. This theory is one of the main reasons why more experienced surgeons are able to

multitask and are less vulnerable to the effects of intra-operative distraction compared with novice surgeons[26]. Consequently, the improvement of cognitive skills prior to a real-life surgical procedure can reduce the cognitive load on the surgeon intra-operatively, which then allows more emphasis to be placed upon learning and performing psychomotor skills. Evidence to support this theory is also present in work by Kohlz-Gatzoulis et al.[140], who implemented cognitive skills training during a surgical skills course for teaching TKA on dry-bone models. The investigators demonstrated that, despite receiving less "hands-on" technical skills training, the surgical performance of surgeons who had undergone extra cognitive skills training was comparable to their control group counterparts at the end of the course. Furthermore, the group receiving cognitive skills training was significantly better at error-detection than the control group - a very favourable outcome given the important role of error detection in surgery[314].

In this study, both methods of cognitive skills training appear to have improved the knowledge and decision-making skills of the participants. However, the magnitude of this improvement was greater in the intervention group than the control group. These results are consistent with those observed in the novice (medical student) population in the earlier development phase of the ViTAL simulator (Chapter 4). The only observed difference (which was expected given their level of surgical experience) was that the orthopaedic trainees in both cohorts obtained better overall marks than the medical students in the pre- and post-training MCQ tests. The results of these tests, combined with the fact that there were no baseline demographic differences in surgical experience, suggest that the most likely explanation for the superior technical performance of the ViTAL Training cohort was the group's cognitive skills. It would have been desirable to

have formally established the base-line technical skills of participants with regards to the OUKA procedure prior to cognitive skills training. However, pilot testing in junior orthopaedic trainees during previous OUKA courses had indicated that the complexity of OUKA surgery, combined with the participants' complete lack of experience in performing this procedure, meant that it was almost impossible for junior trainees to progress very far. Thus, pre-training evaluations would have been unlikely to yield any meaningful results, as the trainees could not have carried out the procedure without a high level of supervision and instruction.

An interesting observation, which deserves further analysis, is the fact that the overall average OATSS scores of the trainees in this study (irrespective of training modality) were better than the *in vivo* OATSS scores of trainees performing real-life OUKA cases during the validation of the OATSS scoring system (Chapter 2). This can be explained by several factors. Firstly, real-life OUKA procedures are more difficult than those conducted on anatomical dry-model bones due to factors such as tissue dissection, intra-operative bleeding, and variations in patients' knee ligament tensions. Secondly, as the participants in the current study were operating on inanimate bones, they were less likely to experience a significant amount of surgical "stress" compared to those who were operating on real patients. Performing real-life surgical procedures would have had a detrimental effect on their OATSS scores, as increased surgical stress is known to impair intra-operative surgical performance[315,316]. Finally, despite the fact that the trainees in the OATSS validation study were more experienced than those in this study (the OATSS score validation study participants had performed OUKAs under supervision), participants in the current study underwent a formal period of intensive cognitive skills training that focused on the current task. This was not the case for the

trainees in the OATSS score validation study – a factor that may have placed them at a disadvantage.

One of the advantages of ViTAL-based cognitive skills training is that it enables trainees to enhance their cognitive skills in a self-directed manner, away from the time-constraints of the real surgical environment. This is an attractive option at a time when service-driven healthcare systems are trying to reduce training programme budgets and use surgical instructors more efficiently. Although there is good evidence to support the effectiveness of computer-based training for teaching basic surgical skills, compared with didactic or video-based teaching[317-319], these studies have primarily focused on basic tasks. The ViTAL simulator addresses a more complex and technically demanding orthopaedic procedure, so the potential educational gains for the trainee are greater. More importantly, these improvements in surgical learning and performance can potentially have a major impact on patient safety by reducing the number of technical and non-technical errors during OUKA surgery.

5.4.2. Limitations

As mentioned previously, evaluation of the base-line technical skills of participants during simulated OUKA procedures on anatomical dry-bone models would have been desirable but was not possible due to their level of surgical experience. Ensuring the homogeneity of the sample population during surgical skills studies can be challenging, and this was particularly difficult given the limited pool of orthopaedic trainees who met the study's eligibility criteria. Despite these problems, the surgical logbook data and pre-training test results of the two cohorts indicated that they were comparable. Another limitation was that

only one blinded assessor evaluated technical performance following the training period. However, this individual was a consultant knee surgeon with fellowship training in OUKA surgery, who also had previous experience of using the OATSS scoring system. In addition, there was high inter-rater reliability between the assessments of this evaluator and an unblinded individual. This not only supports the study findings but also provides further evidence for the reliability of the OATSS scoring system. Finally, the current findings cannot be extrapolated to suggest the transfer validity of ViTAL-based cognitive skills training in the real operating room. However, given the importance of ensuring patient safety, they are a necessary step in achieving the ultimate goal, which is to improve intra-operative performance during OUKA.

5.4.3. Summary

This randomised controlled study has demonstrated that cognitive skills training using the ViTAL simulator can be more effective than standard training modalities for improving surgical performance during simulated OUKA. The results suggest that improving cognitive skills alone can have a profound effect on performance during technically challenging surgical procedures. Methods to help surgeons deal effectively with the cognitive load of complex surgical procedures prior to stepping into a real operating room should become a routine component of surgical training. The successful development of cognitive simulators to serve this purpose could have a profound impact on surgical education and patient care. In the next chapter, the potential role of other novel technological tools, such as Patient Specific Instrumentation systems, which have recently been developed to improve intraoperative performance during OUKA will be explored further.

6. RANDOMISED CONTROLLED TRIAL TO ASSESS THE IMPACT OF PATIENT-SPECIFIC INSTRUMENTATION ON SURGICAL ACCURACY DURING THE OXFORD UNICOMPARTMENTAL KNEE ARTHROPLASTY PROCEDURE

6.1. INTRODUCTION

The work described in previous chapters has focussed on using novel educational tools to enhance training and assessment methods for OUKA surgery. In addition to these techniques, a number of other important modalities have been employed to improve intra-operative performance during this technically challenging surgical procedure. Consequently, the potential benefits of these modalities warrant further investigation. They are primarily based on technological innovations that are used intra-operatively to improve surgical accuracy and component positioning. Traditionally, this has involved the use of computer navigation and robotic systems, with studies demonstrating that both these technologies can result in more accurate component positioning during UKA[194-200]. However, there is still a lack of long-term functional outcome data to support their routine use. Furthermore, factors such as cost, time constraints, and tracker pin site-related problems (e.g. fracture, haematoma formation, nerve palsy and infection), have prevented their wide-scale adoption[201-206]. A new and alternative approach, which has recently received significant attention, has been the introduction of patient-specific instrumentation (PSI) for knee arthroplasty[320-322]. This technology involves the use of pre-operative imaging (MRI or CT) and rapid prototyping technology to produce PSI that fits precisely on a given patient's

femur and tibia. The PSI systems consist of two broad categories. One category consists of patient-specific "pinning" guides that replace conventional alignment instrumentation and enable the placement of pins over which the standard cutting blocks can be placed. The other consists of actual custom "cutting" guides through which direct saw cuts can be made. The process starts with the acquisition of preoperative MRI or CT scans of the patient's knee, in addition to specific slices through the hip and ankle joint, in order to determine limb alignment. The images are segmented and a three-dimensional model of the knee joint surface is reconstructed. Using the computer work-package and virtual templates, the optimum size and position of the components is calculated according to the surgeon's plan and relative to the patient's own anatomy. The plastic cutting jigs that are manufactured have location features on their internal surfaces that correspond to the specific patient's anatomy. During surgery, the jigs are positioned on the patient's bones with an exact anatomical fit (Figure 6-1). Such systems can potentially simplify and improve the accuracy of knee arthroplasty by enabling surgeons to determine the size of bone resections, implant sizes, implant positioning and implant rotation preoperatively. If these aims can be achieved successfully and with relative ease, then such a system will be of great value to those aiming to improve their learning and performance in knee arthroplasty. Furthermore, by removing the need for intra-medullary instrumentation, they may reduce intra-operative blood loss and the risk of fat embolism. Finally, the potential simplification of implantation and the consequent reduction of operative time, together with a reduction in the inventory of standard instruments, can improve operating room efficiency and reduce the overall cost of surgery. A number of companies have now developed PSI guides for their respective

implants. The PSI system that has been developed for the Oxford UKA is called the "Signature" system (Signature™ Personalized Patient Care System; Biomet Inc., Warsaw, IN).

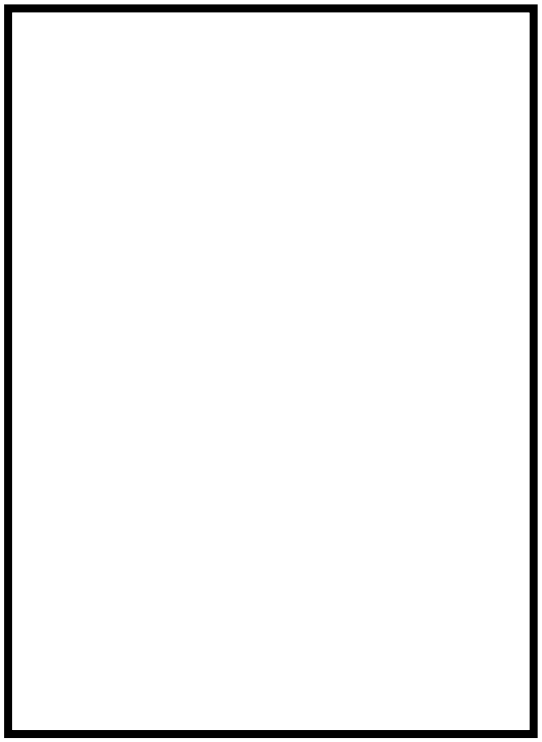

Figure 6-1. Illustration of intra-operative application of the Signature PSI (Biomet Inc.) to the knee joint (Image removed due to copyright licence)

A key goal of enhancing surgical performance in UKA surgery is the development of new instrument systems that can simplify and improve the accuracy of surgical implantation. Ideally, instrumentation should minimise patient morbidity, be easy for surgeons to use and lead to reproducible implantation of prosthesis, whilst also reducing instrument inventory and cost. This type of instrumentation is particularly important for UKA as some knee surgeons perform very few UKAs (compared with TKAs) and so gain less experience in their use. Joint Registry data suggests that a lack of experience with UKA may increase long-term failure rates[192]. Furthermore, previous studies have demonstrated that the accuracy with

which the OUKA prosthesis is implanted is very important in determining the outcome of the procedure[254,323]. Recent case-series and laboratory-based experiments have demonstrated that PSI technology can improve component positioning and alignment during UKA surgery[324-327]. However, no prospective randomised studies have so far been published that evaluate the use of this technology during UKA surgery.

6.1.1. Aim:

To run a prospective randomised controlled trial to compare the accuracy of implantation of an MRI-based patient-specific instrumentation system with conventional instrumentation for Oxford UKA in a high volume centre.

6.2. METHODS

6.2.1. Study design and patients

This RCT was conducted over a twelve-month period at the Nuffield Orthopaedic Centre (NOC). Ethical approval was sought and gained by the National Research Ethics Service committee (REC reference: 11/H0605/1**)** and the NOC review board. A previous pilot study, involving 10 patients, had assessed the safety and feasibility of implanting the Phase 3 mobile bearing OUKA using its manufacturer's PSI system (Signature™ Personalized Patient Care System; Biomet Inc, Warsaw, IN). The four expert OUKA surgeons (DWM, CAFD, AJP, WFMJ) who performed the procedures in the pilot study also performed the procedures in the current study. Patients who were being placed on the waiting list for a medial OUKA, and met the entry criteria for the trial, were invited to participate in the study.

Inclusion criteria were standard for OUKA and as follows:

- Both cruciate ligaments were functionally intact.
- Full thickness cartilage in the lateral compartment.
- Correctable varus deformity.
- Flexion deformity was less than 15°.
- Full thickness cartilage loss in the medial compartment.

Exclusion criteria were as follows:

- Patients for whom a MRI was contra-indicated.
- All forms of inflammatory arthritis.
- Medial or lateral subluxation, or posterior tibial, bone loss suggesting the ACL was damaged.
- If any of the inclusion criteria were not met.

The patients were provided with an information leaflet inviting them to take part in the study. A member of the research team (AA) subsequently contacted the patients in order to determine if they would agree to take part in the study. The patient pathway is presented in Figure 6-2. Patients who wished to take part in the study were consented in the pre-admission clinic and randomised to either the Signature Instrumentation (SI), or Conventional Instrumentation (CI) Group. The SI Group consisted of 23 knees and the CI Group consisted of 22 knees (see sample size calculation below). Randomisation was performed using sealed opaque envelopes. Concealment from the operating surgeon and the patient was

not possible owing to the need for the surgeon to confirm the PSI plans, and the SI Group patients having to undergo an MRI scan preoperatively. Those entered into the SI Group were subsequently recalled for a dedicated MRI scan using the specific protocol outlined by the manufacturers of the Signature system, in order to plan the development of the PSI.

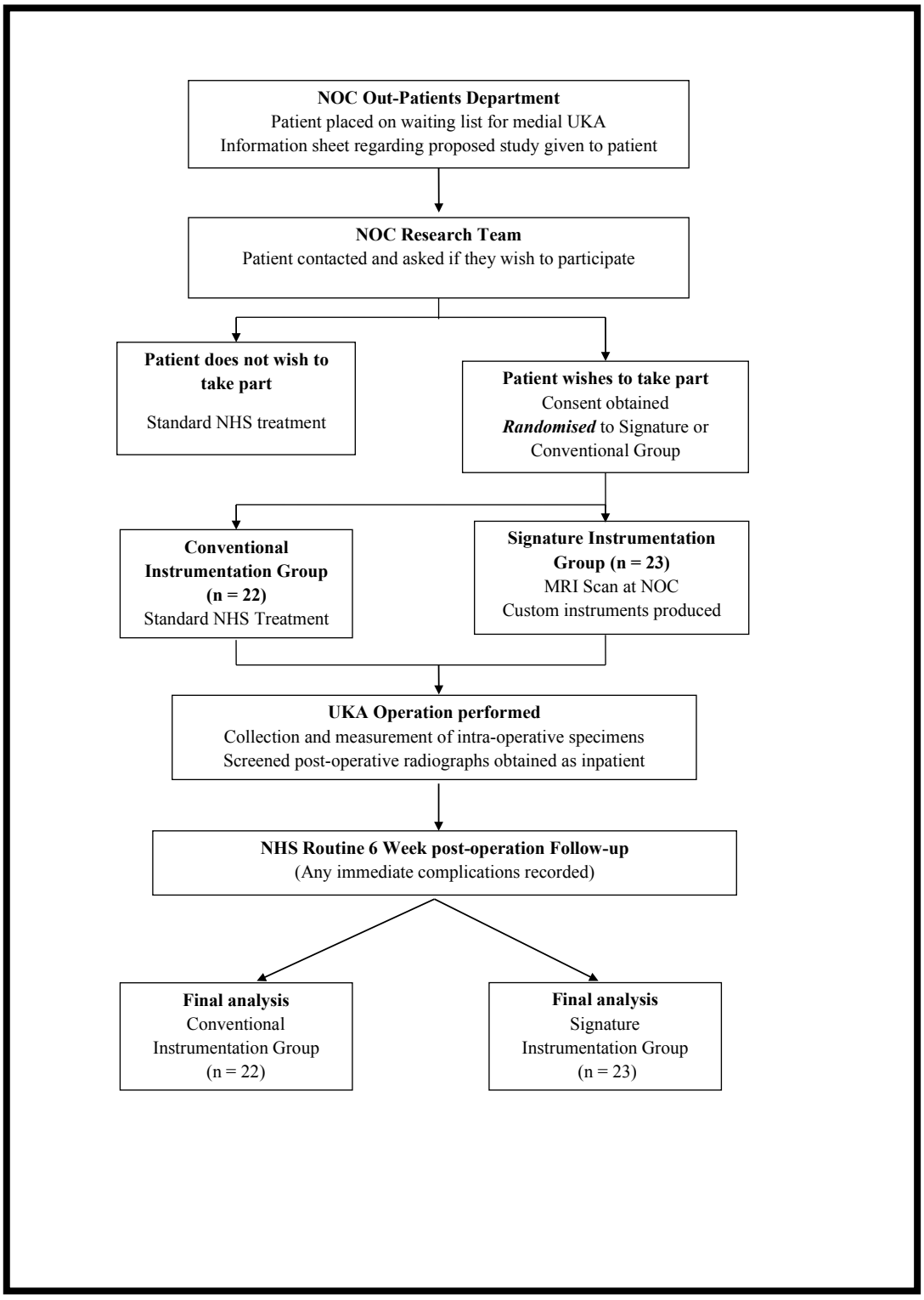

Figure 6-2. Flow of patients through the Signature UKA study.

This preoperative imaging was then sent via a secure server to Biomet Inc., who then forwarded it to Materialise (Leuven, Belgium). Materialise then uploaded the imaging to a proprietary software planner. A preliminary preoperative surgical

plan indicating the prosthesis size, positioning, alignment, and the proposed bone resection levels was electronically uploaded onto the Signature planning platform for online review by the surgical team (Figure 6-3). The default preoperative plans were accepted unless gross errors were present.

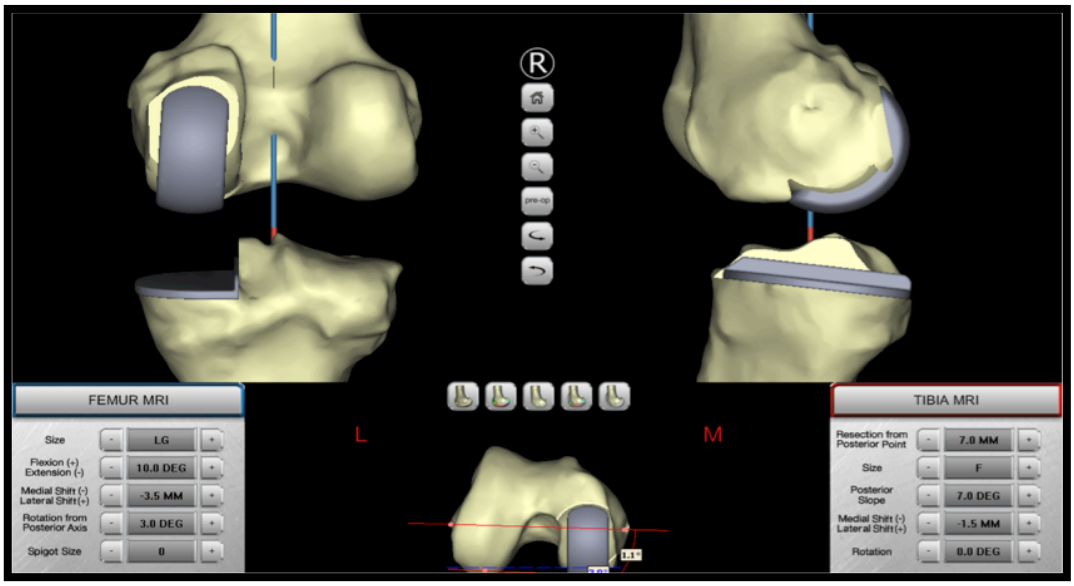

Figure 6-3. Screenshot of the Signature PSI planning software.

Once the final plans were approved, patient-specific cutting blocks were manufactured and sent for sterilisation in preparation for the day of surgery (Figure 6-4).



Figure 6-4. The Signature patient-specific guides and accompanying bone models. Left image is the tibia and right image is the femur.

All procedures were performed under general anaesthetic and a high thigh tourniquet. All patients received a Phase III mobile bearing medial Oxford UKA via a minimally invasive approach by the four aforementioned knee surgeons. The only difference between the two groups was use of PSI. The use of PSI requires the femur to be prepared first as opposed to the tibia. Intra-operatively, drill holes were made in the bone, through predetermined holes in the PSI jig (Figure 6-5). Cutting blocks were inserted in the holes and the bone cuts were made without the use of any intra or extra-medullary instrumentation. The knee was balanced with the mill and completed in the standard fashion.

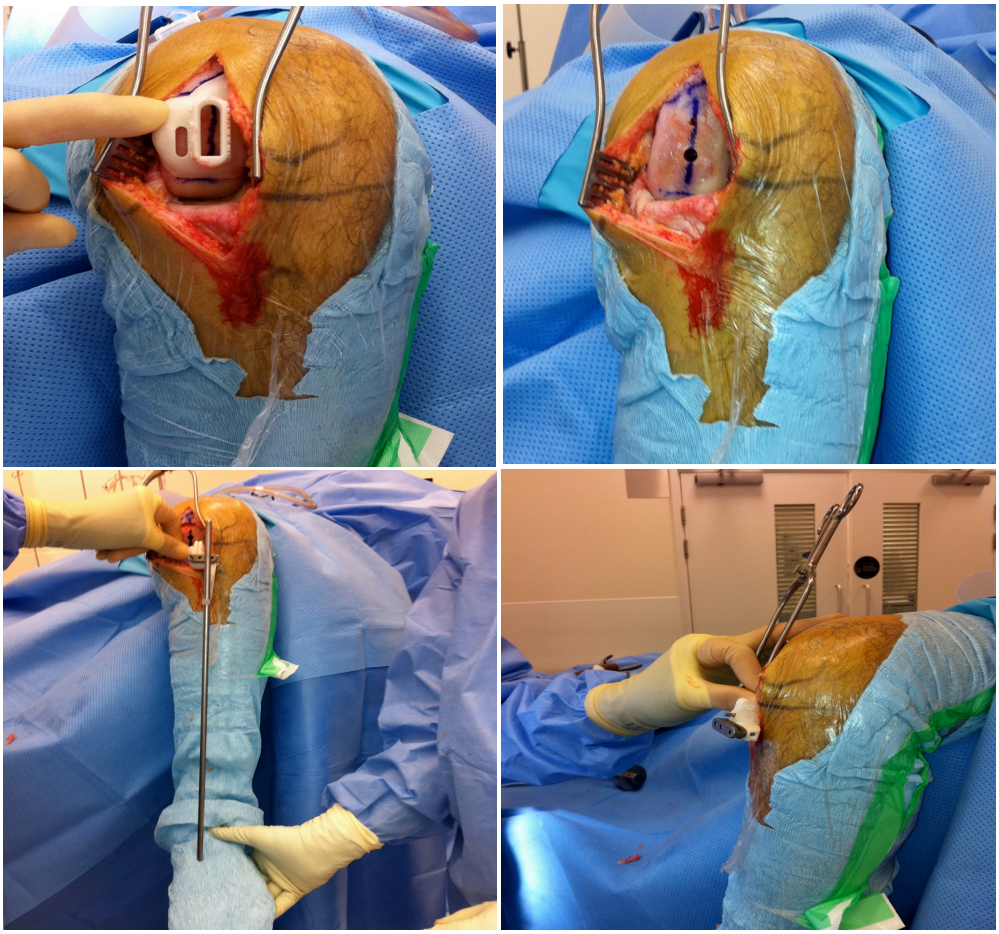

Figure 6-5. Intra-operative use of Signature PSI. Top two images demonstrate their use on the femoral side. Bottom two images demonstrate use on the tibial side.

The Signature PSI for one of the cases could not be used at all, as the guides would not register onto the patient's normal anatomy. For two further cases, the tibial guides could not be used for similar reasons. Conventional instrumentation was, therefore, used for these cases but they were analysed as part of the "SI Group" based on the "intention to treat principle"[328,329]. Operative time was recorded as the "knife to skin" to "skin closure" time. Patients were mobilised on the first post-operative day under the supervision of physiotherapists. Post-operatively, the need to have a blood transfusion, and the change in the pre- to post-operative haemoglobin levels were also recorded. In addition, oxygen saturation levels over the first 24 hours post surgery were recorded (at 0, 6, 12 and 24 hours post surgery). Screened anteroposterior (AP) and lateral (LA) post-operative radiographs were performed prior to discharge. As with all patients undergoing knee replacement surgery at the NOC, study patients attended the physiotherapy ward discharge clinic at six weeks post surgery, where any post-operative complications were recorded.

6.2.2. Outcome measures

6.2.2.1. Primary outcome measures

6.2.2.1.1. Intra-operative assessment of surgical accuracy:

i. The need to perform a horizontal "re-cut" of the tibia.

ii. The size of the meniscal bearing inserted (with optimum bearing size considered to be 3 mm or 4 mm).

iii. Tracking of the meniscal bearing at the end of the procedure (measured as the distance from the bearing to the metal upright of the tibial component in flexion and extension).

iv. Correspondence between the *implanted* and *planned* component sizes.

v. Measurement of resected bone cuts compared to the preoperative plan. This was measured for thickness using engineering callipers (with adjustment for saw cut thickness).

- The central part of the resected femoral posterior facet (mm).
- The central part of the posterior tibial bone cut (mm).

6.2.2.1.2. Post-operative radiological assessment of component positioning:

This was performed on screened AP and LA post-operative radiographs, according to the radiographic parameters proposed by the developers of the Phase 3 OUKA who have provided "optimum" ranges for each parameter[165]. These described the alignment of the femoral and tibial components (Figure 6-6), and also the fit of the tibial component (in terms of overhang or underhang):

i. **Femoral component varus/valgus angle** on the AP radiograph [optimum position: <10° varus to <10° valgus].

ii. **Femoral component flexion/extension angle** on the LA radiograph [optimum position: 15° flexion to <0° extension].

iii. **Tibial component varus/valgus angle** on the AP radiograph [optimum position: <5° varus to <5° valgus].

iv. **Tibial component posterior/inferior tilt angle** on the LA radiograph [baseline value 7° posterior tilt recorded as 0° with optimum position: within +/-5° of the 0° baseline].

v. **Tibial component medial fit** on the AP radiograph [optimum position: Flush or <2 mm overhang].

vi. **Tibial component posterior fit** on the LA radiograph [optimum position: Flush or <2 mm overhang].

vii. **Tibial component anterior fit** on the LA radiograph [optimum position: Flush or <5 mm underhang].

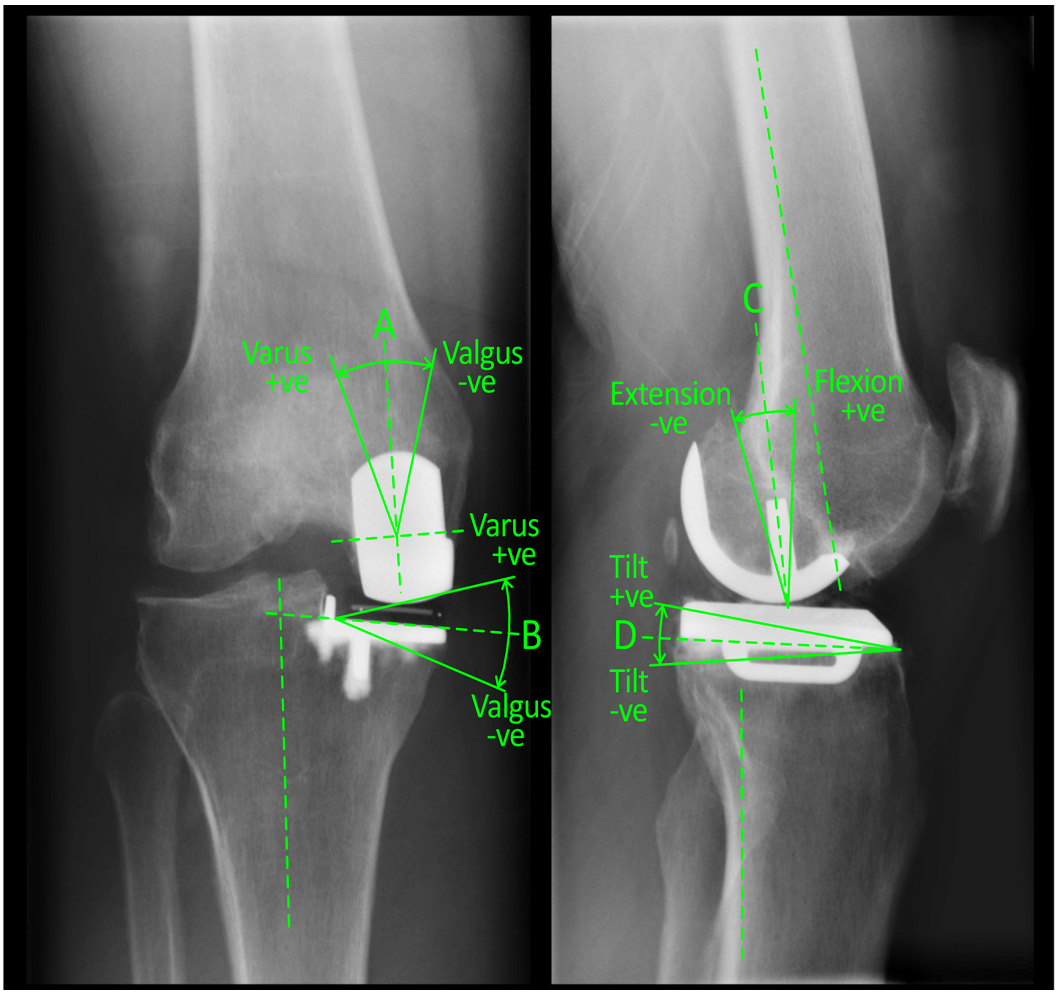

Figure 6-6. Radiographic assessment of component alignment (A = Femoral component varus/valgus angle; B = Tibial component varus/valgus angle; C = Femoral component flexion/extension angle; D = Tibial component poster/inferior tilt angle).

The radiographic parameters were blindly assessed by an independent reviewer (TK), who was not involved in the study, using custom image processing software written and developed using Matlab (v.7.0, The MathsWork Inc., MA, USA). One of the investigators (AA) then blindly measured the radiographic parameters for 20 randomly selected knees twice on two separate days, in order to determine intra- and inter-observer reliability.

6.2.2.2. Secondary outcome measures

The secondary outcome measures were used to further evaluate the efficiency and safety profile of the PSI system. These were:

i. Operative time.

ii. Change in haemoglobin levels post surgery and the need for blood transfusion.

iii. Oxygen saturation measurements over the first 24 hours post surgery, and acute episodes of post-operative confusion (which would indicate the possibility of fat embolism).

iv. Overall impression of the surgeon about PSI used in each case. This was classed as “Good”, “Acceptable”, “Poor”, or “Failure”.

6.2.3. Statistical analysis

Target sample size:

The sample size was calculated from a previous study that used similar radiological assessments to compare minimally invasive OUKAs, performed using conventional instrumentation, with those performed using computer navigation[330].

In this study, the results were normally distributed and the standard deviation of the tibia varus/valgus angle for the control (conventional instrumentation) group was 3.6°. Assuming that the minimum clinically important difference between the control and experimental group means is 3°, the standard mean difference would be 0.8. Using Altman's nomogram, a standard mean difference of 0.8, power of 0.8 and significance of 0.05 results in approximately 44 patients (22 in each group) being required.

The Shapiro-Wilk test showed that all data, except "Bearing alignment", were normally distributed. Parametric tests were applied to normally distributed data, which are presented as means and standard deviations (SD). Where non-parametric tests were applied, data are presented as medians and interquartile ranges (IQRs). The independent group *t*-test was used to compare group demographics, resected bone cut thickness, radiographic parameters, drop in post-operative haemoglobin, and operative time. The Pearson chi-square test was used to compare "radiographic parameter outliers" and "Bearing sizes". The Mann-Whitney *U* test was used to compare "Bearing alignment" in flexion and extension. The inter- and intraobserver reliability of the radiographic parameters was assessed using the intraclass correlation coefficient (ICC). All data were analysed with SPSS software. A p-value < 0.05 was considered statistically significant.

6.3. RESULTS

6.3.1. Participant demographics

As evident in Table 6-1, the groups were equally matched in terms of age, sex, American Society of Anaesthesiologists (ASA) grade, and Body Mass Index (BMI).

Demographic	Signature Instrumentation (n=23)	Conventional Instrumentation (n=22)	p value
Mean Age (SD) years	66.9 (7.2)	68.2 (10.2)	0.621
Sex (M:F)	10 : 13	13 : 9	0.310
Median ASA (Range)	2 (1-3)	2 (1-3)	0.510
Mean Body Mass Index (SD)	29.8 (4.7)	31.8 (5.1)	0.196

Table 6-1. Summary of the study subjects' demographics.

6.3.2. Intra-operative assessment of surgical accuracy

6.3.2.1. Requirement for a tibial plateau "re-cut":

Three cases in the SI group required a horizontal "re-cut" of the tibia compared with two in the CI Group. This difference was not statistically significant (Pearson chi-square value = 0.178, p = 0.673).

6.3.2.2. Bearing size:

All but one of the CI group had a size 3 mm or 4 mm (i.e. optimum size) meniscal bearing inserted compared with only 12 in the SI group (Figure 6-7). This difference was statistically significant (Pearson chi-square value = 10.771, p = 0.001).

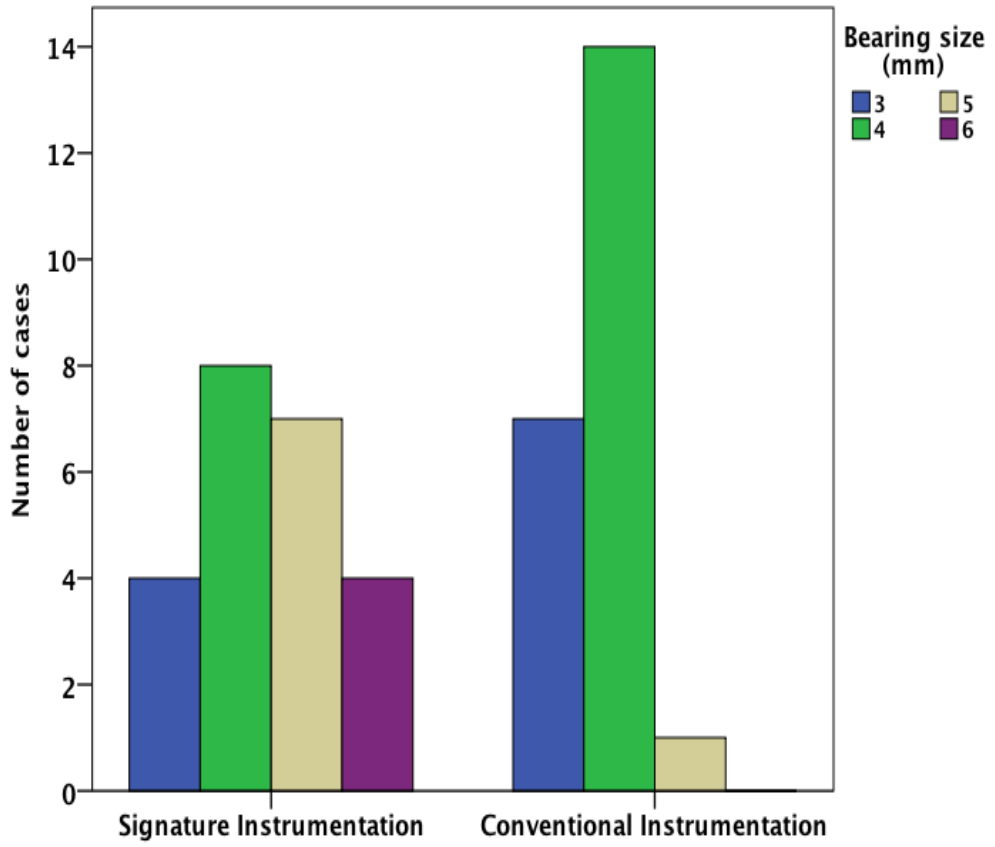

Figure 6-7. Bar chart demonstrating the differing bearing sizes implanted in each group.

6.3.2.3. Bearing tracking:

With regards to tracking of the meniscal bearing (in both flexion and extension), there were no significant differences between the SI and CI groups. The median "bearing to the metal upright of the tibial component wall distance" in flexion was 1 mm (IQR = 1 - 2) for the SI group and 1 mm (IQR = 1 - 1.5) for the CI group. The median distance in extension was 2 mm (IQR = 1 – 2.5) for the SI group and 1 mm (IQR = 1 - 2) for the CI group.

6.3.2.4. Comparison of the implanted and planned component sizes:

When analysing the component sizes that were planned by the PSI system, all but two of the femoral sizes implanted corresponded to the sizes that were planned preoperatively. On the other hand, only eleven of the tibial components implanted corresponded to the sizes that were planned preoperatively (Table 6-2).

	Number of knees in the PSI group (%)
Cases where implanted femoral component = planned femoral component	21 (91 %)
Cases where implanted tibial component = planned tibial component	11 (48 %)

Table 6-2. Comparison between the component sizes implanted and those planned.

6.3.2.5. Thickness of resected bone:

Figure 6-8 demonstrates the difference between the thickness of "planned" and "actual" bone resections for the posterior femoral facets and the tibial plateaus.

Posterior femoral facet: The mean difference for the thickness of resected bone cuts in the SI Group (compared with the estimated bone cuts in the preoperative plan) was 0.1 mm *under*-resection (SD = 0.86), compared with 0.7 mm *under*-resection (SD = 0.68) for the CI group. This difference was statistically significant ($p = 0.01$).

Tibial plateau: The mean difference for the thickness of resected bone cuts in the SI Group (compared with the estimated bone cuts in the preoperative plan) was 0.6 mm *over*-resection (SD = 1.21), compared with *no over/under resection* (SD = 1) for the CI group. This difference was not statistically significant ($p = 0.084$).

It must be noted that during 4 of the 23 SI cases, the +2 mm tibial shim was used to prevent over-resection of the tibia. Thus, the calculated "actual" bone resections for these cases are the measurements that would have been obtained if the Signature PSI plans were followed by the surgeon. However, the shim was used due to the operating surgeon's concern that if the tibial resection had been performed at the level suggested by the PSI guide, the resulting tibial plateau resection would have been unacceptably large.

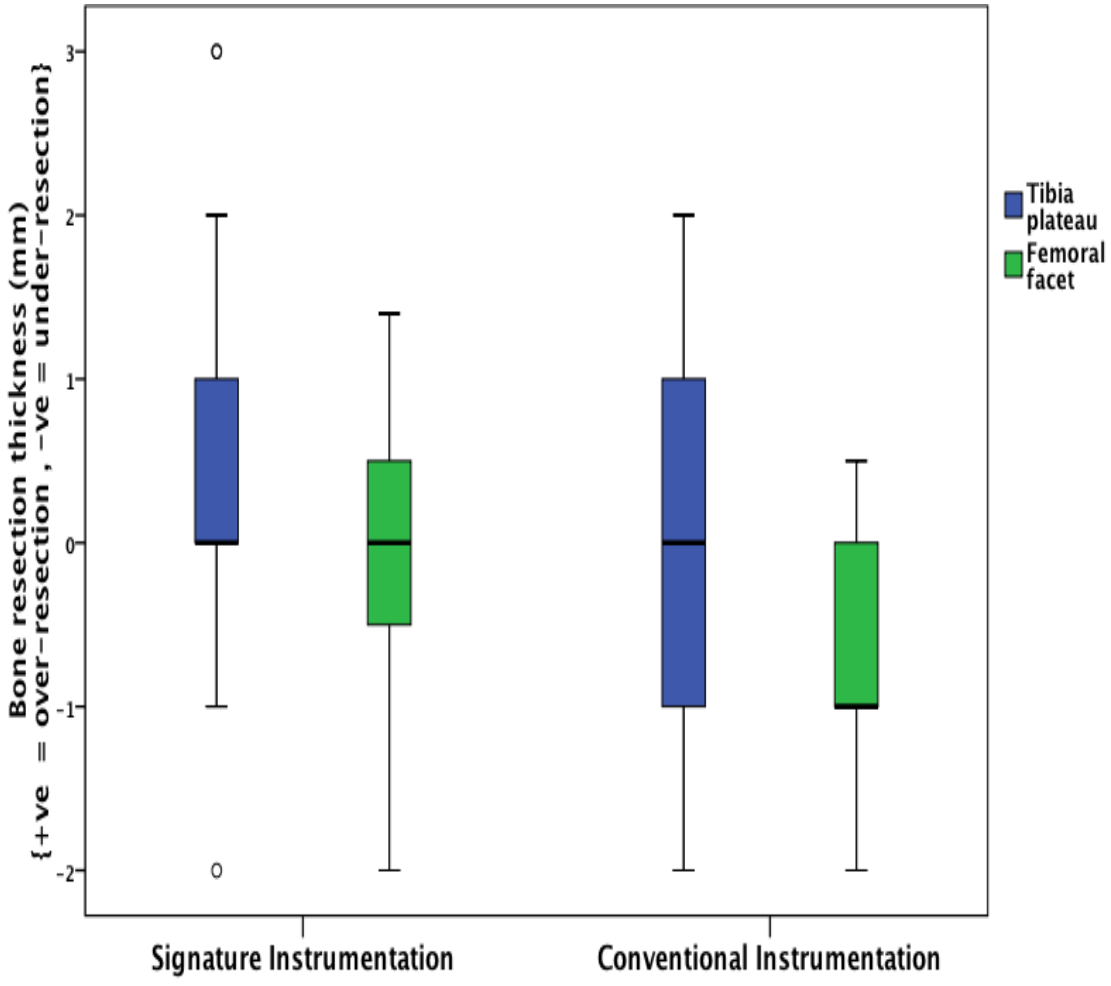

Figure 6-8. Thickness of resected bone expressed as the difference between the "planned" and "actual" resection (the box represents the IQR, the bold line in the box represents the median, the whiskers represent the 2.5 and 97.5 percentiles, and circles are outliers).

6.3.3. Radiological assessment of component positioning

The intra-and inter-observer ICCs for all the measured radiographic parameters was > 0.80 indicating high reliability for these measurements. The proportion of cases that were positioned outside the optimum ranges (i.e. outliers) are summarised in Table 6-3. There were no significant differences between the numbers of outliers within the two groups.

Radiographic Parameter	Signature Instrumentation (n=23)	Conventional Instrumentation (n=22)	p value
	No. of outliers [%]	No. of outliers [%]	
Femur: Varus / Valgus angle	1 (varus) [4%]	None	0.323
Femur: Flexion / Extension angle	None	1 (flexed) [5%]	0.301
Tibia: Varus / Valgus angle	5 (All varus) [22%]	6 (All varus) [27%]	0.666
Tibia: Posterior tilt angle	2 (1 superior, 1 inferior) [9%]	5 (All superior) [3%]	0.194
Tibia: Medial fit	None	1 (underhang) [5%]	0.301
Tibia: Anterior fit	None	None	NA
Tibia: Posterior fit	None	1 (overhang) [5%]	0.301

Table 6-3. Proportion of radiographic parameter outliers in the Signature Instrumentation and Conventional Instrumentation groups (comments in brackets indicate the direction of the outliers).

The mean values of the radiographic parameters for the SI and CI groups are presented Table 6-4. There were no significant statistical differences in component positioning between the two groups, except the "posterior tilt" of the tibial component. The tibial components of the SI group had less posterior tilt (this is expressed as "superior").

Radiographic Parameter	Signature Instrumentation	Conventional Instrumentation	p value (Means)
	Mean (SD)	Mean (SD)	
Femur: Varus / Valgus angle	0.9° varus (4.0)	1.8° varus (3.0)	0.465
Femur: Flexion / Extension angle	9.1° flexion (3.0)	8.8° flexion (4.8)	0.828
Tibia: Varus / Valgus angle	3.5° varus (2.9)	4.0° varus (2.1)	0.585
Tibia: Posterior tilt angle	1.8° superior (2.8)	3.7° superior (2.1)	****0.029**
Tibia: Medial fit	0 mm (1.0)	1 mm underhang (1.3)	0.05
Tibia: Anterior fit	1 mm underhang (0.9)	1 mm underhang (0.7)	0.287
Tibia: Posterior fit	0 mm (0.8)	0 mm (0.9)	0.686

Table 6-4. Summary of the radiographic parameters in the Signature Instrumentation and Conventional Instrumentation groups (denotes statistically significant difference; SD = Standard Deviation).**

6.3.4. Secondary outcome measures

There were no subjects in the study who required a blood transfusion, and there were no significant differences between the two groups in terms of the mean drop in haemoglobin levels post surgery (Table 6-5). None of the subjects in the study suffered from acute post-operative confusion or a drop in oxygen saturation levels (< 95%) during the post-operative period. However, the mean operative time for the SI group was statistically higher than that of the CI group. There were no reported post-operative complications in either group at the six-week follow-up visit.

Secondary outcome measures	Signature Instrumentation (n=23)	Conventional Instrumentation (n=22)	p value
Mean drop in post-operative Hb (SD)	1.7 g/dl (0.8)	1.8 g/dl (0.6)	0.520
Number of patients who had a drop in oxygen saturation levels (<95%) over 24 hours	0	0	-
Mean Operative Time (SD)	75.3 minutes (9.9)	63.5 minutes (11.3)	**0.001****

Table 6-5. Summary of the secondary outcome parameters. (denotes statistically significant difference)**

6.3.5. Surgeons' overall impression of the PSI system

Table 6-6 summarises the surgeon's overall impression of the PSI system for each case. As previously stated, the Signature PSI for one of the cases could not be used as the guides would not register onto the patient's normal anatomy. Hence the PSI for this case was rated as a "Failure". In a further two cases, the tibial guides could not be used for similar reasons. However, the operating surgeons provided their opinions on the femoral guides, as they were satisfied with the fit of the guides on the native femur. One of these cases was rated as "Good" and the other as "Acceptable". From the remaining twenty cases, four guides were rated as "Good", nine guides as "Acceptable", and seven guides as "Poor".

Surgeons' overall impression of the PSI system	Number of cases
Good	5
Acceptable	10
Poor	7
Failure	1

Table 6-6: Summary of the surgeons' overall impression of the PSI system.

6.4. DISCUSSION

6.4.1. Implications of the study

The overall results of this randomised trial demonstrate that (in the hands of expert UKA surgeons) the Signature PSI system is not more accurate than conventional instrumentation with regards to component positioning and alignment of OUKA. Overall accuracy of femoral component positioning was comparable to conventional instrumentation, but the accuracy of tibial component positioning was poor. Therefore, in its current form, this PSI system is not accurate enough or feasible to enable improvement of surgical performance and training in OUKA.

Analysis of the primary intraoperative outcome measures demonstrates that a significantly lower number of cases in the SI group had the "optimum" bearing size inserted. This suggests inaccuracy of the PSI system in terms of planning the tibial plateau resection level. This drawback is further highlighted when considering the difference between the thickness of the "planned" and "actual" bone resections. Although, initially, the average results indicate that there was very little difference between "planned" and "actual" tibial plateau resections, a number of planned PSI resection levels were dangerously large and required the use of the +2 mm shim in order to prevent over-resection. Further evidence of the inaccuracy of the planning algorithms for the tibia is evident when analysing the preoperative planning of component sizes by the PSI system. A large proportion of the planned tibial component sizes were incorrect. However, there appeared to be no significant difference in the final tracking of the meniscal bearings between the two groups, suggesting that the position and direction of the vertical tibial cut was correct. Despite this, it is clear that the PSI system did not *improve* surgical

performance during OUKA and that it may have even been dangerous in the hands of less experienced surgeons such as trainees. These individuals may not have recognised some of the inaccuracies of the PSI jigs during the procedure and would have proceeded to make the resections based on those recommended by the PSI system. Such errors would have resulted in dangerously large bone resections.

The other important primary outcome measures were the post-operative radiographic parameters. Radiographic assessment of component positioning demonstrated that there were equivalent numbers of "outliers" in the two groups, meaning that acceptable alignment and positioning of the prostheses could be achieved using either system. Achieving the correct degree of varus/valgus angulation and posterior tibial slope appeared to be the most challenging task in both groups, as these parameters had the highest proportion of outliers. When comparing the average values for the radiographic parameters, there is some indication that PSI improved the accuracy for achieving the optimum posterior tibial slope angle. However, the actual clinical impact of this measured improvement (which was very small) remains unclear.

The secondary outcome measures demonstrated that both groups had a satisfactory safety profile (in terms of blood loss and risk of fat embolism) and that there were no post-operative complications. The unexpected finding of an increased operative time observed in the SI group could possibly be explained by two factors. The first is the learning curve for this new technology, and the second is the fact that the surgeons intentionally devoted some extra time to evaluating the PSI guides intra-operatively so that they could provide meaningful feedback to the PSI development engineers (with the ultimate goal of further improving the technology).

To date, there have only been a handful of studies that have evaluated the use of PSI technology for UKA surgery. There are no published randomised studies comparing the outcome of using PSI guides with that of conventional instrumentation during UKA surgery. The use of patient-specific *implants* together with PSI has been previously described in case-series[324,325] which used radiographic parameters to confirm the accuracy of component alignment and positioning. Although the use of patient-specific *implants* may be of value for managing complex cases where there is gross deformity or bone loss, the routine use of this technology during everyday practice is not currently feasible or cost-effective. In a recent laboratory-based study using dry bone models, the accuracy of PSI technology for another UKA implant (Uniglide, Corin, Cirencester, UK) was compared to a robot-assisted system and also to conventional instrumentation[327]. The investigators used 3D laser scanning technology to evaluate the accuracy of implant positioning (comparing them with preoperative plans). The results demonstrated that the accuracy of PSI was significantly higher than conventional instrumentation and equivalent to the robot-assisted system. A recent case-series by Volpi et al.[326] has evaluated the accuracy of a PSI system for implanting another UKA prosthesis (Zimmer Unicompartmental High Flex Knee System; Zimmer Inc., Warsaw IN, USA). The investigators demonstrated (using radiological assessment) that the PSI system was "highly accurate in reproducing what the surgeon had planned preoperatively in terms of mechanical axis, tibial slope and component fit to bone".

PSI has been used more frequently for performing TKA than UKA, hence there is a larger body of literature available to evaluate its potential during TKA[320-322]. Nevertheless, very few investigators have reported intra-operative findings and

events (e.g. thickness of bone cuts) as a means of evaluating the accuracy of PSI. In one of the earliest studies evaluating this technology, Spencer et al.[331] reported the incidence of tibial plateau "re-cuts" and the thickness of intra-operative bone resections with favourable measurements supporting the accuracy of PSI. However, they did not state whether the resection measurements for the PSI group differed from those in the preoperative plans. One of the most important technical considerations when performing OUKA is to resect the correct amount of bone from the medial tibial plateau. If the bone resection is too large, there is a danger of damaging the deep fibres of the medial collateral ligament. It was therefore deemed important to reflect this by reporting the meniscal bearing sizes and the bone resection measurements. Although there appeared to be no major issues with the tracking of the meniscal bearings in either group, there was evidence of lower accuracy for bone resections within the SI group. The significantly higher number of the meniscal bearings that were not the "optimum" thickness within the SI group reflects this problem. In addition, further analysis of the SI group demonstrates that a +2 mm shim had to be used in four (17%) of the cases to prevent dangerously large resections of the tibial plateau. Finally, the planning software for this PSI system appeared to be more accurate for the femoral component than the tibial component. This is evident by the fact that less than only half of the planned tibial component sizes corresponded to the sizes that were ultimately implanted. The accuracy of the software-based planning algorithms for this PSI system therefore requires further improvement and re-evaluation by its developers.

The potential intra-operative inaccuracies of PSI systems used for TKA were also highlighted in a randomised controlled study by Victor et al.[332]. In this study, the

investigators used intra-operative computer navigation to validate the accuracy of the PSI jigs of four different companies. The PSI jigs had to be modified in 28% of cases or abandoned altogether in 22% of cases due to intra-operative evidence of gross inaccuracies in alignment and component positioning. The authors also carried out a comprehensive post-operative evaluation of three-planar alignment of the prostheses using plane radiographs and CT scans. Their results demonstrated that PSI guides did not improve the accuracy of alignment in TKA. The overall conclusions that can be drawn from this study are that the PSI systems evaluated during the trial could not improve surgical performance and were unsuitable as a training tool for inexperienced surgeons.

Studies evaluating the effect of PSI on TKA alignment have predominantly used post-operative radiological assessment to determine frontal plane (coronal) alignment. Many of these studies demonstrated that PSI guides offered a more accurate means of achieving a neutral mechanical axis for TKAs, and that they reduced the number of outliers[333,334]. However, these positive findings have not been universal and some studies have found no differences in mechanical axis alignment between PSI and conventional instrumentation[335-337]. The radiological assessment of component alignment and positioning performed in the current study was based on the recommendations of the Oxford UKA design manual and current available literature on these parameters[253,254]. Although no difference was found in the number of outliers between the SI and CI groups, the use of PSI resulted in a marginal improvement in achieving the optimum posterior tibial slope and the optimum medial fit of the tibial component. The clinical relevance of these minor improvements is not yet clear.

Although theoretically plausible, previous studies have failed to show that use of PSI can result in a significant reduction in blood loss or the risk of fat embolism[334,338]. Similarly, it was not possible to show any difference in terms of these potential benefits, as there were no differences in blood loss and none of the patients showed evidence of suffering from fat emboli. Larger studies may be able to demonstrate these potential advantages of using PSI.

6.4.2. Limitations

This study had several limitations. Firstly, the surgeons performing the cases were highly experienced OUKA surgeons with extensive previous experience of using the conventional instrumentation. These experts are therefore likely to have a very low number (if any) of surgical outliers when using the conventional instrumentation. This potential bias is further amplified by the surgical learning curve associated with the use of the PSI. A concerted effort was made by the surgeons to address this issue by familiarising themselves with the Signature instrumentation in a previous pilot study. Nevertheless, the previous experience of the surgeons with the conventional instrumentation is very likely to have favoured the CI group and hence it would have been difficult to demonstrate more superior surgical accuracy using PSI. Furthermore, the high accuracy and low number of outliers observed in the PSI group may not be as applicable to cases performed by less experienced surgeons. Second, the radiological assessment was based on coronal and sagittal alignment. Further evaluation of component rotation using CT scans would have been desirable but was limited by resources. Nevertheless, the radiographic parameters were those that were recommended by the designers of the OUKA prosthesis and previous studies have validated their use for radiological assessment of OUKA[253,254]. Third, the current study does not evaluate possible

differences in the long-term clinical outcome between the two groups. It will ultimately be the long-term functional outcomes and the risk of revision surgery that will determine the efficacy and cost-effectiveness of PSI technology. For this reason, it has been decided that the current study will be part of a larger long-term trial (with appropriate ethical and institutional approval) comparing the clinical outcome between these two groups of patients.

6.4.3. Summary

The results of this study demonstrate that high-volume OUKA surgeons can achieve similar results (in terms of component positioning and alignment) when using PSI guides or conventional instrumentation to implant the OUKA prosthesis. However, detailed intra-operative assessment of surgical accuracy of the tibial guides indicate that in its current form, the PSI system for OUKA has significant drawbacks and requires further improvement. Until these problems have been addressed, PSI for the OUKA should only be used by surgeons experienced in the use of this prosthesis. The current PSI system is therefore not only unsuitable for training purposes, but it could potentially result in further intra-operative errors if used by inexperienced surgeons (e.g. orthopaedic trainees). Even if the technology is improved, it is likely that there will still be some difficulties on the tibial side relating to PSI guide engagement and resection height, so enhancement of training and assessment methods, which have been outlined in previous chapters of this Thesis, will remain the safest and most feasible way of improving surgical performance.

7. DISCUSSION

7.1. SUMMARY OF MAIN FINDINGS AND CONCLUSIONS

It is evident from the literature review that there have been major changes in the design and provision of surgical education over the last two decades. The underlying reasons for these changes are multifactorial. There has been a dramatic reduction in the time available to train surgeons of the future due to doctors' work-hour legislations. This has meant that it is no longer feasible for surgical training to be delivered using the old "Apprenticeship" model, and it is necessary to explore more efficient ways of teaching and assessing surgical skills. In addition, ethical considerations, together with public and peer scrutiny mean that it is not acceptable for surgeons to "practice" on patients. Furthermore, surgical training programmes must demonstrate that surgeons of the future have the necessary skills and attributes to perform their duties safely. Another challenge facing trainee surgeons is the continually increasing number of complex procedures with their accompanying surgical 'learning curve'. Surgical errors are more likely to occur when individuals are beginning to learn a new and complex procedure. A number of potential solutions have, therefore, been proposed for combating this important phenomenon. They include methods to improve the teaching and assessment of surgical procedures, in addition to using new technologies to assist surgeons intra-operatively. The literature review (Chapter 1) highlighted the fact that very few studies had evaluated the application of these methods for commonly performed and technically challenging procedures in orthopaedic surgery. The OUKA was

highlighted as an appropriate example that could be used to further explore these issues. The thesis therefore set out to answer the following questions:

1. Can assessment of surgical performance during the OUKA procedure be improved?

2. Can novel technologies, such as VR-based simulators, improve learning of the OUKA procedure?

3. Can novel technologies, such as patient-specific instrumentation, be used to improve surgical performance during the OUKA procedure?

Given that formative assessment and constructive feedback is an integral component of the learning process, it was deemed necessary to develop a procedure-specific assessment scale for evaluating surgical performance during OUKA. This was addressed in a stepwise process in Chapter 2. In order to determine the content of the assessment scale, the consensus-based Delphi methodology was used (Chapter 2 – Study A) to establish the key surgical steps of the OUKA procedure. Thirty-four expert UKA surgeons from the UK and around the world participated in the Delphi process and completed the necessary survey. This enabled the construction of an 18-point procedure-specific assessment scale, which was named the "Oxford Arthroplasty Technical Skill Score (OATSS). The reliability and validity of OATSS when used in the real operating theatre was determined in Chapter 2 Studies B and C, respectively. In Study B, two trained assessors rated the performance of orthopaedic trainees during 14 OUKA procedures using the OATSS system. The inter-observer reliability was excellent with an ICC of 0.96 for the overall score and a weighted kappa score of 0.71 for the subsections. Study C subsequently compared the performance of expert

OUKA surgeons and orthopaedic trainees during 26 procedures to determine the construct validity of OATSS. The significant difference ($p<0.0005$) between the two cohorts (mean OATSS was 87.7/90 for experts and 63.2/90 for trainees) confirmed the construct validity of this procedure-specific assessment scale.

Other indirect methods for assessing performance during OUKA, and its relationship to clinical outcome, were explored by means of post-operative radiological evaluation. Chapter 3 Study A used another Delphi process to determine the radiological parameters that were considered important for evaluation following OUKA surgery. Thirty UKA experts from around the world and the UK completed the Delphi process that resulted in the construction of a radiological assessment system consisting of 23 parameters. Chapter 3 Study B proceeded to evaluate the clinical relevance of this radiological assessment system. An attempt was made to correlate the risk of a poor functional outcome or prosthesis failure with the radiological assessment system. Several parameters were identified as having a strong influence on clinical outcome, including tibial prosthesis overhang, meniscal bearing misalignment, and poor cementation of the tibial component. Overall, however, it was not possible to correlate the single radiological score to implant failure.

A cognitive simulator, named the Virtual Reality Tool for Arthroplasty Learning (ViTAL), for teaching the OUKA procedure was developed and validated in Chapter 4. The content of ViTAL was partly evidence-based and partly based on the results of the Delphi survey in Chapter 2. A usability study was carried out which consisted of 21 participants with differing backgrounds and levels of surgical expertise. The participants' feedback after using ViTAL was used in order to make further improvements to the simulator. The effectiveness of ViTAL in

teaching the cognitive skills required to perform OUKA was then tested in a randomised study involving 25 medical students. Both cohorts used standard learning material, but the ViTAL cohort also used the cognitive simulator to learn about OUKA. Pre- and post-training MCQ tests were used to determine the participants' knowledge. The results showed that, although both cohorts learned significantly, the ViTAL group performed significantly better ($p<0.0001$) in the post-training MCQ (median score of 85% compared to 53% in the 'standard' training cohort).

It was important to determine whether or not the improvement in cognitive skills that can be achieved by using the ViTAL simulator actually has any effect on surgical performance. Therefore, another randomised study, Chapter 5, was performed in a surgical skills laboratory using synthetic (phantom) knee models. Sixteen orthopaedic trainees who had never performed the OUKA procedure were randomised to undergo either ViTAL or standard training. Knowledge testing consisted of pre- and post-training MCQs, whilst technical skill testing required the participants to perform the OUKA procedure on the synthetic models. The previously validated OATSS system was used to rate surgical performance. Using the ViTAL simulator resulted in better attainment of cognitive knowledge as well as better surgical performance. The mean MCQ score for the ViTAL cohort was 97.5% compared to 87.5% in the standard training cohort ($p<0.05$). The median OATSS score for the ViTAL cohort was 79.6/90 compared to 71.8/90 in the standard training cohort ($p<0.05$).

The value of PSI technology in improving surgical performance during OUKA was evaluated in Chapter 6. There have been no published RCTs in the literature that have evaluated the effectiveness of PSI during UKA surgery, and so it was

felt that the role of this technology needed to be explored further. High-volume surgeons who used either conventional instrumentation or the OUKA PSI system performed forty-five OUKA procedures. Comparisons were made in terms of component alignment, positioning, and accuracy of boney resections. In addition, recording intra-operative and post-operative complications also assessed the safety profile of the two methods. Although, in general, similar component alignment and positioning was achieved using the two systems, the tibial guides for the PSI system were less accurate resulting in over-resection of the tibial plateau in some cases. In addition, the PSI system could not be used in three (out of the 23) cases due to concerns regarding its accuracy and registration onto the native anatomy. There were no immediate or early complications associated with the two systems. The PSI system did not appear to confer any extra advantage in terms of improving the surgical performance of high-volume OUKA surgeons and, in its current form, should not be recommended for inexperienced surgeons.

7.1.1. Contributions to the field of study

This thesis contributes significantly to the field of study in several ways:

- The Oxford Arthroplasty Technical Skill Score (OATSS) for the assessment of OUKA is the first *procedure-specific rating scale* that has been developed and validated for a complex orthopaedic procedure.

- The ViTAL simulator, which was developed for teaching the cognitive components of the OUKA procedure, is the first of its kind for the OUKA procedure.

- The RCT, which involved orthopaedic trainees performing OUKA in the surgical skills laboratory, is the first study in the orthopaedic literature to demonstrate that simulator-based training focusing purely on improving cognitive skills can improve technical skills during arthroplasty surgery.

- The trial involving Patient-Specific Instrumentation (PSI) is the first RCT that has evaluated the effectiveness of this technology during unicompartmental knee arthroplasty procedures.

The findings in this thesis suggest that a procedure-specific rating scale, developed exclusively for OUKA, is a valid and reliable tool that can be utilised for formative assessment during this procedure. Post-operative radiological assessment alone is not a reliable method of predicting clinical outcome following OUKA, but is a useful quality control method for surgeons to critique their own surgical technique. Cognitive skills play a major role in surgical performance and it has been demonstrated that additional cognitive skills training using the ViTAL simulator can significantly improve surgical performance during simulated OUKA. The role of PSI during the OUKA procedure is not yet clear. In its current form, and in the hands of high volume OUKA surgeons, there is not enough evidence to suggest that it is any more advantageous than conventional instrumentation. Thus, improvement of learning and assessment methods is currently the safest and most feasible means of optimising surgical performance during the OUKA procedure.

7.2. CLINICAL RELEVANCE

The findings of this thesis have several important and highly significant clinical implications for surgeons learning to perform the OUKA procedure:

- The procedure-specific rating scale for OUKA, i.e. the OATSS system, has helped to define the critical steps of this challenging procedure based on international expert consensus. The OATSS system can be used as a formative assessment tool in the real operating theatre to monitor the progress of trainees and provide them with feedback as they learn this challenging procedure. Furthermore, the process used for the development and validation of OATSS can be employed as a 'blueprint' for developing procedure-specific rating scales for other complex orthopaedic procedures.

- There is good expert consensus on what radiological parameters are important for post-operative assessment following OUKA. By critiquing post-operative radiographs using these parameters, surgeons can determine what errors (if any) were made and take necessary steps to improve their surgical technique in the future. However, it is not yet possible to accurately predict the clinical outcome of OUKA procedures using post-operative radiographs alone.

- The ViTAL simulator is an exciting educational tool that provides an effective means of learning about the OUKA procedure. Its use alongside traditional learning modalities should be encouraged. An evidence-based approach, together with usability studies, during simulator development will help improve "buy-in" from the end-user and improve the educational impact of the simulator.

- Improvement of the cognitive skills necessary for this surgical procedure by using an appropriate simulator has the potential to further enhance technical performance by reducing the intra-operative cognitive load placed on the surgeon. Although the transfer validity of ViTAL training has not yet been demonstrated for real life OUKA procedures, the methods employed during the development and validation of ViTAL can potentially be used to develop cognitive simulators for other complex orthopaedic procedures.

- Although PSI technology has the potential to improve surgical accuracy during joint arthroplasty, it does not offer any additional improvement for high-volume OUKA surgeons. Furthermore, it currently has a number of limitations that prohibit its routine use during the OUKA procedure - in particular by inexperienced surgeons.

7.3. RECOMMENDATIONS FOR FUTURE WORK

OUKA is now one of the most commonly used knee replacements in the world, and introducing methods to improve the training and performance of surgeons is important. It is anticipated that the findings of this thesis can be used to further improve the methods by which OUKA is taught and assessed, in order to ultimately result in better surgical performance. A number of important questions still remain unanswered and future research is required to further explore these issues.

The development of OATSS (Chapter 2) has enabled the introduction of a procedure-specific assessment scale for evaluating surgical performance for OUKA. However, there are a number of questions related to the psychometric properties of OATSS that warrant further investigation. Although pilot studies suggest that OATSS is sensitive enough to differentiate between skill levels amongst trainees, the floor and ceiling effects for the OATSS system have not yet been fully determined. This would require a very large pool of participants with a wide range of surgical experience (in particular OUKA surgery). The participants would therefore need to range from junior orthopaedic trainees to expert OUKA surgeons in order to investigate this floor and ceiling effect. Undertaking this type of study in a single institution would be logistically difficult due to the number of trainees that would need to be recruited and so it is recommended that it is conducted as a multi-centre study.

Further evaluation of the OATSS system could be performed by comparing its usability and psychometric properties with those of the OCAP PBA tool, which is currently used in assessing surgical performance during knee arthroplasty. Such a

study would have to include feedback from both the trainees and the assessors after they have used both assessment systems.

The OATSS system could also be utilised longitudinally to assess the learning curve of individual orthopaedic trainees during simulated or real-life OUKA procedures. Determining the characteristics of this learning curve is important, as it will provide further evidence for the minimum number of OUKAs that are necessary to reach expert level performance. Using the OATSS system to monitor learning curves and skill improvement during OUKA would also provide further evidence for its use as a training and assessment tool in the real operating room.

Further clarification of the role of radiological parameters, identified by the Delphi process in Chapter 3, and their influence on clinical outcome is warranted. It may be possible for an appropriately weighted radiological score to predict early implant failure or poor outcome for OUKA. Therefore, exploring the development of this type of radiological assessment system may be possible by conducting a large multi-centre cohort study focusing on the technical causes of OUKA procedures that have failed.

It is important to further investigate the role of the ViTAL cognitive simulator (Chapters 4 and 5) in OUKA training. A cadaveric study comparing the performance of orthopaedic trainees is therefore being planned with the ultimate aim of conducting an RCT to assess the transfer validity of ViTAL training to the real operating theatre. In addition, it is necessary to determine whether any additional modifications to ViTAL will further improve its educational potential. A key issue, which requires further investigation, is whether or not additional improvements in simulator fidelity will have any further beneficial effects on

learning and performance. One method of improving fidelity would involve the addition of haptic software that enables users to experience force-feedback when handling instruments and soft tissues. This would, however, limit the widespread availability and possible use of the simulator by significantly increasing its cost. As a result, this type of modification will need to be balanced against the potential gains in educational value. An alternative approach, which is currently being planned for future implementation, is to enable the use of ViTAL on a mobile app platform with subsequent validation studies amongst orthopaedic trainees.

Finally, despite the drawbacks of the PSI system that was evaluated in Chapter 6, this type of technology is continually moving forward and improvements are being made to the software algorithms responsible for the design and development of PSI jigs. This will undoubtedly increase its reliability and accuracy and would enable further assessment of this technology in large multicentre RCTs with long term follow up using functional outcome scores. More importantly, if proven to be successful, future work evaluating PSI systems must include the evaluation of their role in improving surgical learning and performance amongst trainees and inexperienced UKA surgeons, as they will be the ones who will benefit most from this technology.

APPENDICES

Appendix 1: Comparison of radiographic parameters in Cohort 1

This cohort consisted of Design Surgeon failures and matched Design Surgeon Survivors. Continuous variables are listed as means (standard deviation) and categorical variables are expressed as percentages.

Radiographic Parameter	Design Surgeon Failures (n=13)			Design Surgeon Matched Survivors (n=26)		
Femur						
Varus / Valgus angle	2.7 (6.1)			5.0 (4.8)		
Flexion / Extension angle	-0.4 (5.2)			2.0 (4.5)		
Medial / Lateral Position of implant	Central = 100%	Off centre = 0%		Central = 91%	Off centre = 9%	
Posterior Fit (> 4mm overhang)	No = 83%	Yes = 17%		No = 96%	Yes = 4 %	
Anterior impingement of bone on bearing?	Yes = 8%	Maybe = 69%	No = 23%	Yes = 4%	Maybe = 52%	No = 44%
Tibia						
Varus/Valgus angle	5.0 (4.5)			4.5 (2.7)		
Posterior tilt angle	-2.5 (2.7)			-0.8 (2.7)		
Medial fit (mm)	1.7 (1.7)			0.9 (1.1)		
Anterior fit (mm)	-1.2 (1.8)			-0.9 (2.1)		
Posterior fit (mm)	-0.3 (1.2)			-0.5 (0.8)		
Lateral fit	Flush = 62%	Gap = 38%		Flush = 87%	Gap = 13%	
Meniscal bearing alignment	OK = 69%	Misaligned = 31%		OK = 96%	Misaligned = 4%	
Undercut tibial spines?	No = 62%	Yes = 39%		No = 83%	Yes = 17%	
Vertical cut of tibia	On medial tibial spine = 69%	Med to medial tibial spine = 31%	Into lateral tibial spine = 0%	On medial tibial spine = 66%	Med to medial tibial spine = 34%	Into lateral tibial spine = 0%
Depth of vertical cut	OK = 85%	Deep = 15 %		OK = 91%	Deep = 9%	
Depth of keel cut	OK = 92 %	Deep = 8 %		OK = 100 %	Deep = 0%	
Posterior cortex of tibia	Intact = 100%	Not = 0%		Intact = 100%	Not = 0%	
Implant-cement interfaces						
Femoral cement mantle	Complete = 54%	Incomplete = 46%		Complete = 78%	Incomplete = 22 %	
Tibial cement mantle	Complete = 62%	Incomplete = 38%		Complete = 91%	Incomplete = 9%	
Other						
Posterior osteophytes	No = 69 %	Yes = 31		No = 87%	Yes = 13%	
Cementophytes	No = 54%	Yes = 46 %		No = 65%	Yes = 35%	

Appendix 2: Comparison of radiographic parameters in Cohort 2

This cohort consisted of Non-Design Surgeon failures and their matched Non-Design Surgeon Survivors. Continuous variables are listed as means (standard deviation) and categorical variables are expressed as percentages.

Radiographic Parameter	Non-Design Surgeon Failures (n=34)			Non-Design Surgeon Matched Survivors (n=68)		
Femur						
Varus / Valgus angle	1.1 (7.5)			1.3 (5.1)		
Flexion / Extension angle	1.5 (6.2)			2.5 (6.4)		
Medial / Lateral Position of implant	Central = 71%	Off centre = 29%		Central = 88%	Off centre = 12%	
Posterior Fit (> 4mm overhang)	No = 76%	Yes = 24%		No = 85 %	Yes = 15 %	
Anterior impingement of bone on bearing?	Yes = 55 %	Maybe = 36 %	No = 9%	Yes = 13%	Maybe = 63%	No = 24%
Tibia						
Varus/Valgus angle	4.3 (4.8)			4.6 (3.7)		
Posterior tilt angle	-0.7 (5.2)			-0.7 (3.9)		
Medial fit (mm)	-0.2 (2.6)			0 (1.5)		
Anterior fit (mm)	-1.7 (1.2)			-1.1 (1.4)		
Posterior fit (mm)	-0.9 (2.0)			-0.4 (1.4)		
Lateral fit	Flush = 67 %	Gap = 33%		Flush = 79%	Gap = 21%	
Meniscal bearing alignment	OK = 79%	Misaligned = 21%		OK = 94%	Misaligned = 6%	
Undercut tibial spines?	No = 70 %	Yes = 30%		No = 69%	Yes = 31%	
Vertical cut of tibia	On medial tibial spine = 70 %	Med to medial tibial spine = 27%	Into lateral tibial spine = 3%	On medial tibial spine = 84%	Med to medial tibial spine = 15%	Into lateral tibial spine = 1%
Depth of vertical cut	OK = 94 %	Deep = 6%		OK = 71%	Deep = 29%	
Depth of keel cut	OK = 100 %	Deep = 0%		OK = 97%	Deep = 3%	
Posterior cortex of tibia	Intact = 97%	Not = 3%		Intact = 98 %	Not = 2 %	
Implant-cement interfaces						
Femoral cement mantle	Complete = 30%	Incomplete = 70 %		Complete = 46%	Incomplete = 54%	
Tibial cement mantle	Complete = 58%	Incomplete = 42%		Complete = 85%	Incomplete = 15%	
Other						
Posterior osteophytes	No = 82 %	Yes = 18%		No = 78%	Yes = 22%	
Cementophytes	No = 41%	Yes = 59%		No = 55%	Yes = 45%	

Appendix 3: Comparison of of radiographic parameters in Cohort 3

This cohort consisted of Design Surgeon Top Decile ΔOKS and Bottom Decile ΔOKS groups. Continuous variables are listed as means (standard deviation) and categorical variables are expressed as percentages.

Radiographic Parameter	Design Surgeon Top OKS Decile (n=41)			Design Surgeon Bottom OKS Decile (n=38)		
Femur						
Varus / Valgus angle	3.1 (4.7)			1.1 (3.6)		
Flexion / Extension angle	0.7 (5.3)			1.5 (6.5)		
Medial / Lateral Position of implant	Central = 88%	Off centre = 12%		Central = 95%	Off centre = 5%	
Posterior Fit (> 4mm overhang)	No = 93%	Yes = 7 %		No = 79 %	Yes = 21%	
Anterior impingement of bone on bearing?	Yes = 0%	Maybe = 73%	No= 27%	Yes = 8%	Maybe = 71%	No = 21%
Tibia						
Varus/Valgus angle	3.4 (3.4)			3.9 (4.2)		
Posterior tilt angle	-0.6 (3.4)			-1.6 (3.4)		
Medial fit (mm)	0.2 (0.9)			0.3 (1.6)		
Anterior fit (mm)	-1.1 (1.1)			-0.8 (1.1)		
Posterior fit (mm)	-0.5 (1.1)			-0.1 (2.7)		
Lateral fit	Flush = 93%	Gap = 7%		Flush = 76%	Gap = 24%	
Meniscal bearing alignment	OK = 100%	Misaligned = 0%		OK = 92%	Misaligned = 8%	
Undercut tibial spines?	No = 76%	Yes = 24%		No = 64%	Yes = 36%	
Vertical cut of tibia	On medial tibial spine = 90%	Med to medial tibial spine = 10%	Into lateral tibial spine = 0%	On medial tibial spine = 16%	Med to medial tibial spine = 84%	Into lateral tibial spine = 0%
Depth of vertical cut	OK = 95%	Deep = 5%		OK = 92%	Deep = 8%	
Depth of keel cut	OK = 100%	Deep = 0%		OK = 100%	Deep = 0%	
Posterior cortex of tibia	Intact = 98%	Not = 2%		Intact = 95%	Not = 5%	
Implant-cement interfaces						
Femoral cement mantle	Complete = 79%	Incomplete = 21%		Complete = 66%	Incomplete = 34%	
Tibial cement mantle	Complete = 95%	Incomplete = 5%		Complete = 84%	Incomplete = 16%	
Other						
Posterior osteophytes	No = 78%	Yes = 22%		No = 72%	Yes = 28%	
Cementophytes	No = 66%	Yes = 34%		No = 55%	Yes = 45%	

Appendix 4: ViTAL Usability Study: Participant demographics

User-demographics

For each question, please circle the most appropriate answer

1. Age:

2. Sex (please circle)**:**

Male Female

3. Experience (please circle)**:**

Consultant / OR Scrub Staff / SpR / Core Trainee / Medical Student / Non-Clinician

4. Does your daily work involve regular use of computers (please circle)**?**

Yes No

5. Level of computing knowledge (please tick one or more category)**:**

a) Don't use
b) Email/internet / social networking / entertainment (I-Tunes/YouTube/etc.)
c) Word processing / presentations
d) Data base entry / manipulation
e) Scientific use (statistical analysis / equation solving)
f) Graphics design / video editing / music composition
g) Software development / Programming (e.g. database implementation)

6. How often do you use computer game? (Please tick <u>one</u> category)

a) Never
b) Less than once a month
c) About once a month
d) A couple of times a month
e) Once or twice a week
f) 3 or 4 times a week
g) 5 or 6 times a week
h) At least once a day

7. If you use computer games, what genre of games do you play (please tick one or more category)**?**

a) First person shooter (e.g. Golden eye / Doom / Call of Duty)
b) Role playing / Adventure games (Diablo / Dungeons and Dragons)
c) Action / Platform games (Super Mario / Street Fighter)
d) Sports games (FIFA / NBA)
e) Driving / Aircraft game
f) Puzzle / casual game (Tetris / Angry Bird / Chess)
g) On-line multiplayer game (Second Life / World of Warcraft)
h) Other (please specify)

8. **Previous exposure to VR software (Serious Games):**

Yes No

9. Previous operative experience (Clinicians and OR Staff only):

- **Number of Oxford UKA procedures performed as primary surgeon:**
  - None
  - < 10
  - 10 – 30
  - 30 – 100
  - > 100
- **Number of Oxford UKA procedure assisted:**
  - None
  - < 10
  - 10 – 30
  - 30 – 100
  - > 100

Appendix 5: Genre of video games played by participants.

(*Note the same participant may play multiple genres of game)

Genre of games played	Number of participants*
First person shooter (e.g. Golden eye / Doom / Call of Duty)	11
Role playing / Adventure games (Diablo / Dungeons and Dragons)	6
Action / Platform games (Super Mario / Street Fighter)	6
Sports games (FIFA / NBA)	7
Driving / Aircraft game	5
Puzzle / casual game (Tetris / Angry Bird / Chess)	13
On-line multiplayer game (Second Life / World of Warcraft)	2
Other	2

Appendix 6: MCQs used for evaluation of the ViTAL Cognitive simulator

Please select the one correct answer (stem) for each question:

1. **Prior to setting up the femoral drill guide assembly for drilling the femoral peg holes, it is necessary to:**

A. Remove the ACL in order to improve exposure and access
B. Insert a 0mm spigot and mill the femoral condyle
C. Insert an intramedullary (IM) rod into the femoral canal at a point situated 1cm anterior to the anteromedial corner of the intercondylar notch
D. Resect the posterior femoral condyle using the special slotted femoral saw guide

2. Towards which anatomical structure should the narrow reciprocating saw blade be pointing when performing the vertical tibial saw cut?

A. The xiphisternum
B. The anterior superior iliac spine / femoral head of the same limb
C. The anterior superior iliac spine of the opposite limb
D. The lateral aspect of the intercondylar notch

3. Which of the following structure must be specifically protected with a retractor during execution of the horizontal tibial cut?

A. Popliteus muscle
B. Medial meniscus
C. Popliteal artery
D. Medial collateral ligament

4. What is the optimum position for drilling the femoral peg holes?

A. The central third of the medial femoral condyle
B. The intercondylar notch of the femur
C. The medial third of the medial femoral condyle
D. The lateral third of the medial femoral condyle

5. Following excision of the posterior facet of the femoral condyle with the saw, the next appropriate step is to:

A. Insert a 0 spigot and mill the femoral condyle
B. Insert a trial femoral component
C. Insert the femoral trimming guide and use the posterior osteophyte chisel to remove any posterior osteophytes
D. Measure the 90° flexion gap

6. Why should the self-retaining retractor be removed when assessing the flexion gaps and ligament balancing the knee?

A. The retractor can impinge on the components and prevent full flexion
B. The retractor artificially diminishes the gaps
C. The retractor can damage the medial collateral ligament
D. The retractor can impede a clear view of the joint

7**. What is the optimum position of the leg whilst the cement is setting on femoral component?**

A. 45° flexion and compression
B. 90° flexion and compression
C. Full extension and compression
D. 90° flexion and no compression

8. Which one of the following is a contra-indication for Oxford Unicompartmental Knee Arthroplasty (OUKA)?

A. Above average Body Mass Index
B. Old age
C. Rheumatoid Arthritis
D. Patellofemoral joint osteoarthritis

9. Following resection of the tibial plateau, the next appropriate step is to:

A. Insert a tibial template and 4 mm feeler gauge to ensure enough bone has been resected
B. Resect the posterior femoral condyle and then check the 90° and 20° flexion gaps
C. Use a tibial template, a 4 mm feeler gauge, and a femoral guide to ensure enough bone has been resected
D. Prepare the tibial plateau keel cut with a saw

10. Attempted excision of osteophytes from which of the following areas should be avoided?

A. Osteophytes on the medial margin of the medial femoral condyle
B. The "anvil" osteophyte situated anterior to the ACL insertion on the tibial plateau
C. Osteophytes in the intercondylar notch of the femur
D. Medial tibial plateau osteophytes beneath the Medial Collateral Ligament

11. What is the optimum position of the leg whilst the cement is setting on tibial component?

A. Full extension and compression
B. 90° flexion and no compression
C. 45° flexion and compression
D. 90° flexion and compression

12. What is the optimum position of the vertical tibial saw cut made with the narrow reciprocating saw?

A. Just medial to the apex of the medial tibial spine
B. Adjacent to the lateral femoral condyle
C. Just lateral to the apex of the lateral tibial spine and the origin of the ACL
D. Between the central and the middle third of the medial tibial plateau

13. Which of the following complications may occur as a consequence of raising the handle of the saw during execution of the vertical tibial saw cut?

A. Stretching of the knee ligaments
B. Damage to the MCL
C. A tibial plateau fracture
D. Damage to the anterior condyle of the femur

14. What size spigot should be used during the first milling of the femoral condyle?

A. "1" spigot
B. "4" spigot
C. Depends on the 90° flexion gap measurement
D. "0" spigot

15. Following joint exposure the next appropriate step is to:

A. Excise any osteophytes which are present
B. Assess the lateral compartment
C. Carry out the horizontal tibial saw cut
D. Improve exposure by releasing the MCL

16. What is a source of bearing impinging ?

A. Anterior and posterior bone from the lateral femoral condyle
B. Anterior bone from the medial tibial plateau
C. Posterior bone from the medial tibial plateau
D. Anterior bone from the medial femoral condyle

17. Following gap balancing and ensuring there is no impingement on the trial meniscal bearing, what step must be undertaken prior to inserting the trial tibial component?

A. The tibial keel slot must prepared using the tibial template and saw
B. The tibial surface must be roughened with a key drill
C. The bone cement should be mixed
D. Medial tibial plateau osteophytes beneath the MCL should be excised

18. After the second milling, which was done with a 3 spigot, the 90° flexion gap measures 5mm and the 20° flexion gap is 4mm. What should be done?

A. The 4 spiggot should be used
B. The 1 spiggot should be used
C. The 1 spiggot should be hammered in until it's collar rests on the condyle
D. The 5 spiggot should be used

19. Which of the following is *NOT* a necessary requirement with regards to equipment set-up prior to drilling the femoral peg holes?

A. The feeler gauge must be touching the vertical wall of the tibial template
B. The slotted femoral saw guide must be placed on the femur
C. The handle of the femoral drill guide must be aligned parallel to the long axis of the tibia
D. The lateral surface of the femoral drill guide must be parallel with the intramedullary (IM) rod.

20. Following drilling of the femoral peg holes, the next appropriate step is to:

A. Insert the slotted femoral guide and excise the posterior facet of the femoral condyle with a saw
B. Insert a 0 spigot and mill the femoral condyle
C. Insert the femoral trimming guide and use the posterior osteophyte chisel to remove any posterior osteophytes
D. Insert a trial femoral component

21. When determining an individual's suitability for OUKA, what information can be obtained from the lateral knee radiograph that might be of relevance?

A. Patellar height
B. Intact lateral compartment
C. Preserved posterior cartilage on the tibia
D. Full thickness loss of cartilage in the medial compartment

22. Which of the following statements is incorrect with regards to patient positioning prior to OUKA?

A. The hip should be flexed > 60°
B. The knee must be free to flex at least 120°
C. The thigh support must not impinge in the popliteal fossa.
D. A special thigh support can be used for positioning the limb

23. Following the first milling of the femoral condyle, the next immediate step is to:

A. Remove the 0 spigot and insert a trial meniscal bearing
B. Remove the 0 spigot and trim the bone from the posterior corners of the femoral condyle and around the spigot hole
C. Remove the 0 spigot, insert the femoral trimming guide, and use the posterior osteophyte chisel to remove any posterior osteophytes
D. Remove the 0 spigot and measure the 20° flexion gap

24. Following the assessment of the knee joint for contra-indications, the next appropriate step is to:

A. Excise any osteophytes present on the intercondylar notch
B. Improve exposure by carrying out a medial release, including release of the MCL
C. Resect the tibial plateau using the horizontal saw
D. Assess the 90° flexion gap

25. Which of the following is considered to be a contra-indication to performing an Oxford Unicompartmental Knee Arthroplasty (OUKA)?

A. Partial cartilage loss on the medial compartment
B. Intact lateral compartment cartilage
C. Functionally intact ACL
D. Partial cartilage loss in the patellofemoral joint

26. During application of the tibial saw guide, the surgeon must ensure that:

A. The guide lies parallel to the long axis of the tibial shaft in both planes
B. The guide is parallel to the anatomical axis of the femur
C. The face of the guide is sitting on the medial femoral condyle
D. The MCL is released in order to allow optimum seating of the guide

27. During application of the tibial saw guide, the surgeon must ensure that:

A. When viewed from the side, the tibial guide is approximately 2cm from the tibial crest
B. The tibial cut at least 5mm below the deepest area of bone erosion
C. The guide is parallel to the anatomical axis of the femur
D. The face of the guide is resting against the medial femoral condyle

28. During the initial stage of the procedure, which of the following would indicate that the OUKA procedure should be abandoned in favour of a total knee replacement?

A. Partial tear of the ACL
B. Large osteophytes on the intercondylar notch of the femur
C. An “anvil” osteophyte anterior the ACL insertion on the tibia
D. Full thickness ulcer on the on the medial side of the lateral condyle

29. What is the optimum position of the 6mm femoral peg hole?

A. The central third of the medial femoral condyle

B. 1cm anterior to the anteromedial corner of the intercondylar notch of the femur
C. Within the central third of intercondylar notch of the femur
D. The medial third of the medial femoral condyle

30. Following resection of the tibial plateau, the next appropriate step is to:

A. Check the 90° and 20° flexion gaps
B. Drill the 6mm femoral peg holes
C. Insert an intramedullary (IM) rod into the femoral canal at a point situated 1cm anterior to the anteromedial corner of the intercondylar notch
D. Insert a tibial template and 4 mm feeler gauge to ensure enough bone has been resected

REFERENCES:

1. **History of the College.** Royal College of Surgeons of Edinburgh: http://www.rcsed.ac.uk/the-college/history.aspx (date last accessed 3 January 2013).

2. **Satava RM, Gallagher AG, Pellegrini CA.** Surgical competence and surgical proficiency: definitions, taxonomy, and metrics. *Journal of the American College of Surgeons* 2003;196-6:933-7.

3. **Sidhu RS, Grober ED, Musselman LJ, Reznick RK.** Assessing competency in surgery: where to begin? *Surgery* 2004;135-1:6-20.

4. **Arora S, Sevdalis N, Suliman I, Athanasiou T, Kneebone R, Darzi A.** What makes a competent surgeon?: experts' and trainees' perceptions of the roles of a surgeon. *American Journal of Surgery* 2009;198-5:726-32.

5. **Bhatti NI, Cummings CW.** Competency in surgical residency training: defining and raising the bar. *Academic Medicine* 2007;82-6:569-73.

6. **The CanMEDS Framework.** Royal College of Physicians and Surgeons of Canada. http://www.royalcollege.ca/portal/page/portal/rc/canmeds/framework (date last accessed 12 March 2013).

7. **Accreditation Council on Graduate Medical Education Common Program Requirements. 2010.** Available at**:** http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/CPRs2013.pdf (date last accessed 12 March 2013).

8. **Cuschieri A, Francis N, Crosby J, Hanna GB.** What do master surgeons think of surgical competence and revalidation? *American Journal of Surgery* 2001;182-2:110-6.

9. **Kirkman MA.** Deliberate practice, domain-specific expertise, and implications for surgical education in current climes. *Journal of Surgical Education* 2013;70-3:309-17.

10. **Alderson D.** Developing expertise in surgery. *Medical Teacher* 2010;32-10:830-6.

11. **Sadideen H, Alvand A, Saadeddin M, Kneebone R.** Surgical experts: born or made? *International Journal of Surgery* 2013;11-9:773-8.

12. **Abernethy B, Poolton JM, Masters RS, Patil NG.** Implications of an expertise model for surgical skills training. *ANZ Journal of Surgery* 2008;78-12:1092-5.

13. **Schaverien MV.** Development of expertise in surgical training. *Journal of Surgical Education* 2010;67-1:37-43.

14. **Ericsson KA.** Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine* 2004;79-10 Suppl:S70-81.

15. **Hatano G, Inagaki K.** Two courses of expertise. In: Stevenson H, Azuma H, Hakuta K, eds. *Child Development and Education in Japan*. New York: W H Freeman 1986:27–36.

16. **Bransford JD, Brown AL, CockingRR.** *How People Learn: Brain, Mind, Experience, and School: Expanded Edition.* Washington, DC: National Academy Press, 2000.

17. **Mylopoulos M, Regehr G.** Cognitive metaphors of expertise and knowledge: prospects and limitations for medical education. *Medical Education* 2007;41-12:1159-65.

18. **Ericsson KA, Kintsch W.** Long-term working memory. *Psychological Review* 1995;102-2:211-45.

19. **Schmidt HG, Norman GR, Boshuizen HP.** A cognitive perspective on medical expertise: theory and implication. *Academic Medicine* 1990;65-10:611-21.

20. **Schneider W, Shiffrin RM.** Controlled and Automatic Human Information-Processing .1. Detection, Search, and Attention. *Psychological Review* 1977;84-1:1-66.

21. **Shiffrin RM, Schneider W.** Controlled and Automatic Human Information-Processing .2. Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review* 1977;84-2:127-90.

22. **Logan GD.** Automaticity, Resources, and Memory - Theoretical Controversies and Practical Implications. *Human Factors* 1988;30-5:583-98.

23. **Stefanidis D, Scerbo MW, Korndorffer JR, Jr., Scott DJ.** Redefining simulator proficiency using automaticity theory. *American Journal of Surgery* 2007;193-4:502-6.

24. **Stefanidis D, Scerbo MW, Sechrist C, Mostafavi A, Heniford BT.** Do novices display automaticity during simulator training? *American Journal of Surgery* 2008;195-2:210-3.

25. **Bann S, Khan M, Datta V.** Surgical skill is predicted by the ability to detect errors. *American Journal of Surgery* 2005;189-4:412-5.

26. **Hsu KE, Man FY, Gizicki RA, Feldman LS, Fried GM.** Experienced surgeons can do more than one thing at a time: effect of distraction on performance of a simple laparoscopic and cognitive task by experienced and novice surgeons. *Surgical Endoscopy* 2008;22-1:196-201.

27. **Schmidt HG, Rikers RM.** How expertise develops in medicine: knowledge encapsulation and illness script formation. *Medical Education* 2007;41-12:1133-9.

28. **Halm EA, Lee C, Chassin MR.** Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Annals of Internal Medicine* 2002;137-6:511-20.

29. **Chowdhury MM, Dagash H, Pierro A.** A systematic review of the impact of volume of surgery and specialization on patient outcome. *British Journal of Surgery* 2007;94-2:145-61.

30. **Nuttall M, van der Meulen J, Phillips N, Sharpin C, Gillatt D, McIntosh G, Emberton M.** A systematic review and critique of the literature relating hospital or surgeon volume to health outcomes for 3 urological cancer procedures. *Journal of Urology* 2004;172-6:2145-52.

31. **McGrath PD, Wennberg DE, Dickens JD, Siewers AE, Lucas FL, Malenka DJ, Kellett MA, Ryan TJ.** Relation between operator and hospital volume and outcomes following percutaneous coronary interventions in the era of the coronary stent. *JAMA : Journal of the American Medical Association* 2000;284-24:3139-44.

32. **Dewey TM, Herbert MA, Ryan WH, Brinkman WT, Smith R, Prince SL, Mack MJ.** Influence of Surgeon Volume on Outcomes With Aortic Valve Replacement. *Annals of Thoracic Surgery* 2012;93-4:1107-13.

33. **Losina E, Barrett J, Mahomed NN, Baron JA, Katz JN.** Early failures of total hip replacement - Effects of surgeon volume. *Arthritis and Rheumatism* 2004;50-4:1338-43.

34. **Shervin N, Rubash HE, Katz JN.** Orthopaedic procedure volume and patient outcomes - A systematic literature review. *Clinical Orthopaedics and Related Research* 2007-457:35-41.

35. **Critchley RJ, Baker PN, Deehan DJ.** Does surgical volume affect outcome after primary and revision knee arthroplasty? A systematic review of the literature. *Knee* 2012;19-5:513-8.

36. **Lau RL, Perruccio AV, Gandhi R, Mahomed NN.** The role of surgeon volume on patient outcome in total knee arthroplasty: a systematic review of the literature. *BMC Musculoskeletal Disorders* 2012;13.

37. **Baker P, Jameson S, Critchley R, Reed M, Gregg P, Deehan D.** Center and surgeon volume influence the revision rate following unicondylar knee replacement: an analysis of 23,400 medial cemented unicondylar knee replacements. *J Bone Joint Surg Am* 2013;95-8:702-9.

38. **Ericsson KA.** Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med* 2008;15-11:988-94.

39. **Ericsson KA, Lehmann AC.** Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology* 1996;47:273-305.

40. **Carpenter SK, Cepeda NJ, Rohrer D, Kang SHK, Pashler H.** Using Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction. *Educational Psychology Review* 2012;24-3:369-78.

41. **Donovan JJ, Radosevich DJ.** A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology* 1999;84-5:795-805.
42. **Brydges R, Carnahan H, Backstein D, Dubrowski A.** Application of motor learning principles to complex surgical tasks: Searching for the optimal practice schedule. *Journal of Motor Behavior* 2007;39-1:40-8.
43. **Mackay S, Morgan P, Datta V, Chang A, Darzi A.** Practice distribution in procedural skills training - A randomized controlled trial. *Surgical Endoscopy and Other Interventional Techniques* 2002;16-6:957-61.
44. **Moulton CAE, Dubrowski A, MacRae H, Graham B, Grober E, Reznick R.** Teaching surgical skills: What kind of practice makes perfect? A randomized, controlled trial. *Annals of Surgery* 2006;244-3:400-9.
45. **Mitchell EL, Lee DY, Sevdalis N, Partsafas AW, Landry GJ, Liem TK, Moneta GL.** Evaluation of distributed practice schedules on retention of a newly acquired surgical skill: a randomized trial. *American Journal of Surgery* 2011;201-1:31-9.
46. **Luft AR, Buitrago MM.** Stages of motor skill learning. *Molecular Neurobiology* 2005;32-3:205-16.
47. **Hall JC.** Imagery practice and the development of surgical skills. *American Journal of Surgery* 2002;184-5:465-70.
48. **Battig WF.** Intratask interference as a source of facilitation in transfer and retention. In: Voss JF, ed. *Topics in learning and performance*. New York: Academic Press, 1972:131-59.
49. **Shea JB, Morgan RL.** Contextual Interference Effects on the Acquisition, Retention, and Transfer of a Motor Skill. *Journal of Experimental Psychology-Human Learning and Memory* 1979;5-2:179-87.
50. **Magill RA, Hall KG.** A Review of the Contextual Interference Effect in Motor Skill Acquisition. *Human Movement Science* 1990;9-3-5:241-89.
51. **Albaret JM, Thon B.** Differential effects of task complexity on contextual interference in a drawing task. *Acta Psychologica* 1998;100-1-2:9-24.
52. **Dubrowski A, Backstein D, Abughaduma R, Leidl D, Carnahan H.** The influence of practice schedules in the learning of a complex bone-plating surgical task. *American Journal of Surgery* 2005;190-3:359-63.
53. **Rasmussen J.** Skills, Rules, and Knowledge - Signals, Signs, and Symbols, and Other Distinctions in Human-Performance Models. *IEEE Transactions on Systems Man and Cybernetics* 1983;13-3:257-66.
54. **Wentink M, Stassen LP, Alwayn I, Hosman RJ, Stassen HG.** Rasmussen's model of human behavior in laparoscopy training. *Surgical Endoscopy* 2003;17-8:1241-6.

55. **Pugh CM, Santacaterina S, DaRosa DA, Clark RE.** Intra-operative decision making: more than meets the eye. *Journal of Biomedical Informatics* 2011;44-3:486-96.
56. **Fitts PM, Posner MI.** *Human Performance.* Belmont, CA: Brooks/Cole, 1967.
57. **Newell KM.** Knowledge of Results and Motor Learning. *Journal of Motor Behavior* 1974;6-4:235-44.
58. **Kopta JA.** The development of motor skills in orthopaedic education. *Clinical Orthopaedics and Related Research* 1971;75:80-5.
59. **Kopta JA.** An approach to the evaluation of operative skills. *Surgery* 1971;70-2:297-303.
60. **Collins A.** Cognitive Apprenticeship. *Cambridge Handbook of the Learning Sciences* 2006:47-60.
61. **Wanzel KR, Ward M, Reznick RK.** Teaching the surgical craft: From selection to certification. *Curr Probl Surg* 2002;39-6:573-659.
62. **Subramonian K, Muir G.** The 'learning curve' in surgery: what is it, how do we measure it and can we influence it? *BJU International* 2004;93-9:1173-4.
63. **Khan N, Abboudi H, Khan MS, Dasgupta P, Ahmed K.** Measuring the surgical 'learning curve': methods, variables and competency. *BJU International* 2014;113-3:504-8.
64. **Guttmann D, Graham RD, MacLennan MJ, Lubowitz JH.** Arthroscopic rotator cuff repair: the learning curve. *Arthroscopy* 2005;21-4:394-400.
65. **Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J.** Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *American Journal of Surgery* 2003;185-2:146-9.
66. **Alvand A, Auplish S, Gill H, Rees J.** Innate arthroscopic skills in medical students and variation in learning curves. *J Bone Joint Surg Am* 2011;93-19:e115(1-9).
67. **Alvand A, Auplish S, Khan T, Gill HS, Rees JL.** Identifying orthopaedic surgeons of the future: the inability of some medical students to achieve competence in basic arthroscopic tasks despite training: a randomised study. *J Bone Joint Surg Br* 2011;93-12:1586-91.
68. **Jackson WF, Khan T, Alvand A, Al-Ali S, Gill HS, Price AJ, Rees JL.** Learning and retaining simulated arthroscopic meniscal repair skills. *J Bone Joint Surg Am* 2012;94-17:e132.
69. **Pollard TC, Khan T, Price AJ, Gill HS, Glyn-Jones S, Rees JL.** Simulated hip arthroscopy skills: learning curves with the lateral and supine patient positions: a randomized trial. *J Bone Joint Surg Am* 2012;94-10:e68.

70. **Bokhari MB, Patel CB, Ramos-Valadez DI, Ragupathi M, Haas EM.** Learning curve for robotic-assisted laparoscopic colorectal surgery. *Surgical Endoscopy* 2011;25-3:855-60.

71. **Eversbusch A, Grantcharov TP.** Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. *Surgical Endoscopy* 2004;18-10:1514-8.

72. **Zevin B, Aggarwal R, Grantcharov TP.** Simulation-based training and learning curves in laparoscopic Roux-en-Y gastric bypass. *British Journal of Surgery* 2012;99-7:887-95.

73. **Grantcharov TP, Funch-Jensen P.** Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. *American Journal of Surgery* 2009;197-4:447-9.

74. **Ellis H.** *A history of surgery.* London: Greenwich Medical Media, 2001.

75. **Halsted W.** *The training of the surgeon : the annual address in medicine delivered at Yale University, June 27, 1904.* Baltimore, 1904:25 p.

76. **Rodriguez-Paz JM, Kennedy M, Salas E, Wu AW, Sexton JB, Hunt EA, Pronovost PJ.** Beyond "see one, do one, teach one": toward a different training paradigm. *Quality & Safety in Health Care* 2009;18-1:63-8.

77. **Feanny MA, Scott BG, Mattox KL, Hirshberg A.** Impact of the 80-hour work week on resident emergency operative experience. *American Journal of Surgery* 2005;190-6:947-9.

78. **Richards T.** European Working Time Directive. Running out of time. *BMJ* 2009;338:b1507.

79. **Philibert I, Friedmann P, Williams WT.** New requirements for resident duty hours. *JAMA : Journal of the American Medical Association* 2002;288-9:1112-4.

80. **Chikwe J, de Souza AC, Pepper JR.** No time to train the surgeons. *BMJ* 2004;328-7437:418-9.

81. **Antiel RM, Thompson SM, Hafferty FW, James KM, Tilburt JC, Bannon MP, Fischer PR, Farley DR, Reed DA.** Duty hour recommendations and implications for meeting the ACGME core competencies: views of residency directors. *Mayo Clinic Proceedings* 2011;86-3:185-91.

82. **Simpson C, Cottam H, Fitzgerald JE, Giddings CE.** The European working time directive has a negative impact on surgical training in the UK. *The Surgeon* 2011;9-1:56-7.

83. **Crofts TJ, Griffiths JM, Sharma S, Wygrala J, Aitken RJ.** Surgical training: an objective assessment of recent changes for a single health board. *BMJ* 1997;314-7084:891-5.

84. **Bridges M, Diamond DL.** The financial impact of teaching surgical residents in the operating room. *American Journal of Surgery* 1999;177-1:28-32.
85. **Clamp JA, Baiju D, Sr., Copas DP, Hutchinson JW, Rowles JM.** Do Independent Sector Treatment Centres (ISTC) impact on specialist registrar training in primary hip and knee arthroplasty? *Annals of the Royal College of Surgeons of England* 2008;90-6:492-6.
86. **Mason A, Street A, Verzulli R.** Private sector treatment centres are treating less complex patients than the NHS. *Journal of the Royal Society of Medicine* 2010;103-8:322-31.
87. **Health Do.** Modernising Medical Careers: The next steps – The Future Shape of Foundation, Specialist and General Practice Training Programmes. 2004.
88. **Darzi A, Smith S, Taffinder N.** Assessing operative skill. Needs to become more objective. *BMJ* 1999;318-7188:887-8.
89. **Aggarwal R, Darzi A.** Technical-skills training in the 21st century. *New Eng J Med* 2006;355-25:2695-6.
90. **Reznick RK, MacRae H.** Teaching surgical skills--changes in the wind. *New Eng J Med* 2006;355-25:2664-9.
91. **Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB.** Observational tools for assessment of procedural skills: a systematic review. *American Journal of Surgery* 2011;202-4:469-80 e6.
92. **Gallagher AG, Ritter EM, Satava RM.** Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surgical Endoscopy* 2003;17-10:1525-9.
93. **Downing SM.** Reliability: on the reproducibility of assessment data. *Medical Education* 2004;38-9:1006-12.
94. **Reznick R, Regehr G, MacRae H, Martin J, McCulloch W.** Testing technical skill via an innovative "bench station" examination. *Am J Surg* 1997;173-3:226-30.
95. **Regehr G, MacRae H, Reznick RK, Szalay D.** Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine* 1998;73-9:993-7.
96. **Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M.** Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;84-2:273-8.
97. **Mackay S, Datta V, Chang A, Shah J, Kneebone R, Darzi A.** Multiple Objective Measures of Skill (MOMS): a new approach to the assessment of technical ability in surgical trainees. *Ann Surg* 2003;238-2:291-300.
98. **van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J.** Objective assessment of technical surgical skills. *Br J Surg* 2010;97-7:972-87.

99. Intercollegiate Surgical Curriculum Programme (ISCP) Website: https://http://www.iscp.ac.uk/ (Last accessed September 2013).
100. **Memon MA, Brigden D, Subramanya MS, Memon B.** Assessing the surgeon's technical skills: analysis of the available tools. *Academic Medicine* 2010;85-5:869-80.
101. **Insel A, Carofino B, Leger R, Arciero R, Mazzocca AD.** The development of an objective model to assess arthroscopic performance. *J Bone Joint Surg Am* 2009;91-9:2287-95.
102. **Elliott MJ, Caprise PA, Henning AE, Kurtz CA, Sekiya JK.** Diagnostic knee arthroscopy: a pilot study to evaluate surgical skills. *Arthroscopy* 2012;28-2:218-24.
103. **Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, Cosgarea AJ, Harner CD, Garrett WE, Olson T, Warme WJ, Nicandri GT.** The Arthroscopic Surgical Skill Evaluation Tool (ASSET). *American Journal of Sports Medicine* 2013;41-6:1229-37.
104. **Koehler RJ, Nicandri GT.** Using the arthroscopic surgery skill evaluation tool as a pass-fail examination. *J Bone Joint Surg Am* 2013;95-23:e1871-6.
105. **Slade Shantz JA, Leiter JR, Collins JB, MacDonald PB.** Validation of a global assessment of arthroscopic skills in a cadaveric knee model. *Arthroscopy* 2013;29-1:106-12.
106. **Alvand A, Logishetty K, Middleton R, Khan T, Jackson WF, Price AJ, Rees JL.** Validating a global rating scale to monitor individual resident learning curves during arthroscopic knee meniscal repair. *Arthroscopy* 2013;29-5:906-12.
107. **Olson T, Koehler R, Butler A, Amsdell S, Nicandri G.** Is there a valid and reliable assessment of diagnostic knee arthroscopy skill? *Clinical Orthopaedics and Related Research* 2013;471-5:1670-6.
108. **Howells NR, Brinsden MD, Gill RS, Carr AJ, Rees JL.** Motion analysis: a validated method for showing skill levels in arthroscopy. *Arthroscopy* 2008;24-3:335-42.
109. **Datta V, Mackay S, Mandalia M, Darzi A.** The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *Journal of the American College of Surgeons* 2001;193-5:479-85.
110. **Moorthy K, Munz Y, Dosis A, Bello F, Chang A, Darzi A.** Bimodal assessment of laparoscopic suturing skills: construct and concurrent validity. *Surgical Endoscopy* 2004;18-11:1608-12.
111. **Chami G, Ward JW, Phillips R, Sherman KP.** Haptic feedback can provide an objective assessment of arthroscopic skills. *Clinical Orthopaedics and Related Research* 2008;466-4:963-8.

112. **Tashiro Y, Miura H, Nakanishi Y, Okazaki K, Iwamoto Y.** Evaluation of skills in arthroscopic training based on trajectory and force data. *Clinical Orthopaedics and Related Research* 2009;467-2:546-52.

113. **Tuijthof GJ, Horeman T, Schafroth MU, Blankevoort L, Kerkhoffs GM.** Probing forces of menisci: what levels are safe for arthroscopic surgery. *Knee Surgery, Sports Traumatology, Arthroscopy* 2011;19-2:248-54.

114. **Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Newhouse JP, Weiler PC, Hiatt HH.** Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. 1991. *Qual Saf Health Care* 2004;13-2:145-51; discussion 51-2.

115. **Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, Hebert L, Newhouse JP, Weiler PC, Hiatt H.** The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *New Eng J Med* 1991;324-6:377-84.

116. **Neale G, Woloshynowych M, Vincent C.** Exploring the causes of adverse events in NHS hospital practice. *Journal of the Royal Society of Medicine* 2001;94-7:322-30.

117. **Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD.** The Quality in Australian Health Care Study. *Medical Journal of Australia* 1995;163-9:458-71.

118. **Gawande AA, Thomas EJ, Zinner MJ, Brennan TA.** The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery* 1999;126-1:66-75.

119. **Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, Etchells E, Ghali WA, Hebert P, Majumdar SR, O'Beirne M, Palacios-Derflingher L, Reid RJ, Sheps S, Tamblyn R.** The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Canadian Medical Association Journal* 2004;170-11:1678-86.

120. **Leape LL.** Error in medicine. *JAMA : Journal of the American Medical Association* 1994;272-23:1851-7.

121. **Nuland SB.** Mistakes in the operating room--error and responsibility. *New Eng J Med* 2004;351-13:1281-3.

122. **de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA.** The incidence and nature of in-hospital adverse events: a systematic review. *Quality & Safety in Health Care* 2008;17-3:216-23.

123. **Kohn LT, Corrigan J, Donaldson MS.** *To err is human : building a safer health system.* Washington, D.C. ; [Great Britain]: National Academy Press, 2000.

124. **Roberts KH.** Managing High-Reliability Organizations. *California Management Review* 1990;32-4:101-13.

125. **Roberts KH.** Some Characteristics of One Type of High Reliability Organization. *Organization Science* 1990;1-2:160-76.
126. **Gaba DM.** Structural and organizational issues in patient safety: A comparison of health care to other high-hazard industries. *California Management Review* 2000;43-1:83-9.
127. **Zeltser MV, Nash DB.** Approaching the Evidence Basis for Aviation-Derived Teamwork Training in Medicine. *American Journal of Medical Quality* 2010;25-1:13-23.
128. **Kao LS, Thomas EJ.** Navigating towards improved surgical safety using aviation-based strategies. *Journal of Surgical Research* 2008;145-2:327-35.
129. **Smith R.** All changed, changed utterly. British medicine will be transformed by the Bristol case. *BMJ* 1998;316-7149:1917-8.
130. **Treasure T.** Lessons from the Bristol case. More openness--on risks and on individual surgeons' performance. *BMJ* 1998;316-7146:1685-6.
131. **Davies JM.** Painful inquiries: lessons from Winnipeg. *Canadian Medical Association Journal* 2001;165-11:1503-4.
132. **Dunbar JA, Reddy P, Beresford B, Ramsey WP, Lord RS.** In the wake of hospital inquiries: impact on staff and safety. *Medical Journal of Australia* 2007;186-2:80-3.
133. **Association of College and University Examiners., Bloom BS.** *Taxonomy of educational objectives : the classification of educational goals. Handbook 1, Cognitive domain.* London: Longmans, 1956.
134. **Candy PC.** *Self-direction for lifelong learning : a comprehensive guide to theory and practice.* San Francisco: Jossey-Bass, 1991.
135. **Gaba DM.** The future vision of simulation in health care. *Qual Saf Health Care* 2004;13 Suppl 1:i2-10.
136. **Kneebone R, Kidd J, Nestel D, Asvall S, Paraskeva P, Darzi A.** An innovative model for teaching and learning clinical procedures. *Medical Education* 2002;36-7:628-34.
137. **Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL.** Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. *J Bone Joint Surg Br* 2008;90-4:494-9.
138. **Butler A, Olson T, Koehler R, Nicandri G.** Do the skills acquired by novice surgeons using anatomic dry models transfer effectively to the task of diagnostic knee arthroscopy performed on cadaveric specimens? *J Bone Joint Surg Am* 2013;95-3:e15(1-8).
139. **Martin KD, Belmont PJ, Schoenfeld AJ, Todd M, Cameron KL, Owens BD.** Arthroscopic basic task performance in shoulder simulator model correlates with similar task performance in cadavers. *J Bone Joint Surg Am* 2011;93-21:e1271-5.

140. **Kohls-Gatzoulis JA, Regehr G, Hutchison C.** Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. *Can J Surg* 2004;47-4:277-83.

141. **Sonnadara RR, Van Vliet A, Safir O, Alman B, Ferguson P, Kraemer W, Reznick R.** Orthopedic boot camp: examining the effectiveness of an intensive surgical skills course. *Surgery* 2011;149-6:745-9.

142. **Leong JJ, Leff DR, Das A, Aggarwal R, Reilly P, Atkinson HD, Emery RJ, Darzi AW.** Validation of orthopaedic bench models for trauma surgery. *J Bone Joint Surg Br* 2008;90-7:958-65.

143. **Holland JP, Waugh L, Horgan A, Paleri V, Deehan DJ.** Cadaveric hands-on training for surgical specialties: is this back to the future for surgical skills development? *Journal of Surgical Education* 2011;68-2:110-6.

144. **Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM.** Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 2002;236-4:458-63; discussion 63-4.

145. **Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR.** Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *British Journal of Surgery* 2008;95-9:1088-97.

146. **Seymour NE.** VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World Journal of Surgery* 2008;32-2:182-8.

147. **Palter VN, Grantcharov TP.** Virtual reality in surgical skills training. *Surgical Clinics of North America* 2010;90-3:605-17.

148. **Mabrey JD, Gillogly SD, Kasser JR, Sweeney HJ, Zarins B, Mevis H, Garrett WE, Jr., Poss R, Cannon WD.** Virtual reality simulation of arthroscopy of the knee. *Arthroscopy* 2002;18-6:E28.

149. **Cannon WD, Eckhoff DG, Garrett WE, Jr., Hunter RE, Sweeney HJ.** Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. *Clinical Orthopaedics and Related Research* 2006;442:21-9.

150. **Gomoll AH, O'Toole RV, Czarnecki J, Warner JJ.** Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. *American Journal of Sports Medicine* 2007;35-6:883-8.

151. **Gomoll AH, Pappas G, Forsythe B, Warner JJ.** Individual skill progression on a virtual reality simulator for shoulder arthroscopy: a 3-year follow-up study. *American Journal of Sports Medicine* 2008;36-6:1139-42.

152. **Mabrey JD, Reinig KD, Cannon WD.** Virtual reality in orthopaedics: is it a reality? *Clinical Orthopaedics and Related Research* 2010;468-10:2586-91.

153. **Modi CS, Morris G, Mukherjee R.** Computer-simulation training for knee and shoulder arthroscopic surgery. *Arthroscopy* 2010;26-6:832-40.
154. **Aggarwal R, Undre S, Moorthy K, Vincent C, Darzi A.** The simulated operating theatre: comprehensive training for surgical teams. *Quality & Safety in Health Care* 2004;13 Suppl 1:i27-32.
155. **Martel-Pelletier J, Boileau C, Pelletier JP, Roughley PJ.** Cartilage in normal and osteoarthritis conditions. *Best Pract Res Clin Rheumatol* 2008;22-2:351-84.
156. No authors listed. Arthritis Research UK. http://www.arthritisresearchuk.org/arthritis-information/conditions/arthritis/who-gets-it.aspx [Date accessed 22/02/2014].
157. **Chen A, Gupte C, Akhtar K, Smith P, Cobb J.** The Global Economic Cost of Osteoarthritis: How the UK Compares. *Arthritis* 2012;2012:698709.
158. **March LM, Bachmeier CJ.** Economics of osteoarthritis: a global perspective. *Bailliere's Clinical Rheumatology* 1997;11-4:817-34.
159. **Peat G, McCarney R, Croft P.** Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. *Annals of the Rheumatic Diseases* 2001;60-2:91-7.
160. **Kurtz S, Ong K, Lau E, Mowat F, Halpern M.** Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am* 2007;89-4:780-5.
161. No authors listed. 10th Annual Report 2013. National Joint Registry for England and Wales. http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/Reports/10th_annual_report/NJR 10th Annual Report 2013 B.pdf. [Date accessed 05/04/2014]
162. **Hernborg JS, Nilsson BE.** The natural course of untreated osteoarthritis of the knee. *Clinical Orthopaedics and Related Research* 1977-123:130-7.
163. **Ledingham J, Regan M, Jones A, Doherty M.** Radiographic patterns and associations of osteoarthritis of the knee in patients referred to hospital. *Annals of the Rheumatic Diseases* 1993;52-7:520-6.
164. **Wise BL, Niu J, Yang M, Lane NE, Harvey W, Felson DT, Hietpas J, Nevitt M, Sharma L, Torner J, Lewis CE, Zhang Y.** Patterns of compartment involvement in tibiofemoral osteoarthritis in men and women and in whites and African Americans. *Arthritis Care & Research* 2012;64-6:847-52.
165. **Goodfellow J, O'Connor J, Dodd CA, Murray DW.** *Unicompartmental arthroplasty with the Oxford knee.* Oxford ; New York: Oxford University Press, 2006:vi, 194 p.

166. **White SH, Ludkowski PF, Goodfellow JW.** Anteromedial osteoarthritis of the knee. *J Bone Joint Surg Br* 1991;73-4:582-6.
167. **McKeever DC.** The classic: Tibial plateau prosthesis.1960. *Clinical Orthopaedics and Related Research* 2005;440:4-8; discussion 3.
168. **Laskin RS.** Unicompartmental tibiofemoral resurfacing arthroplasty. *J Bone Joint Surg Am* 1978;60-2:182-5.
169. **Insall J, Aglietti P.** A five to seven-year follow-up of unicondylar arthroplasty. *J Bone Joint Surg Am* 1980;62-8:1329-37.
170. **Borus T, Thornhill T.** Unicompartmental knee arthroplasty. *Journal of the American Academy of Orthopaedic Surgeons* 2008;16-1:9-18.
171. **Griffin T, Rowden N, Morgan D, Atkinson R, Woodruff P, Maddern G.** Unicompartmental knee arthroplasty for the treatment of unicompartmental osteoarthritis: a systematic study. *ANZ journal of surgery* 2007;77-4:214-21.
172. **Berend KR, Cushner FD.** *Partial knee arthroplasty : techniques for optimal outcomes.* Philadelphia, Pa. ; London: Saunders, 2011.
173. **Brown NM, Sheth NP, Davis K, Berend ME, Lombardi AV, Berend KR, Della Valle CJ.** Total knee arthroplasty has higher postoperative morbidity than unicompartmental knee arthroplasty: a multicenter analysis. *Journal of Arthroplasty* 2012;27-8 Suppl:86-90.
174. **Lombardi AV, Jr., Berend KR, Walter CA, Aziz-Jacobo J, Cheney NA.** Is recovery faster for mobile-bearing unicompartmental than total knee arthroplasty? *Clinical Orthopaedics and Related Research* 2009;467-6:1450-7.
175. **Hopper GP, Leach WJ.** Participation in sporting activities following knee replacement: total versus unicompartmental. *Knee Surgery, Sports Traumatology, Arthroscopy* 2008;16-10:973-9.
176. **Price AJ, Rees JL, Beard DJ, Gill RH, Dodd CA, Murray DM.** Sagittal plane kinematics of a mobile-bearing unicompartmental knee arthroplasty at 10 years: a comparative in vivo fluoroscopic analysis. *Journal of Arthroplasty* 2004;19-5:590-7.
177. **Robertsson O, Borgquist L, Knutson K, Lewold S, Lidgren L.** Use of unicompartmental instead of tricompartmental prostheses for unicompartmental arthrosis in the knee is a cost-effective alternative. 15,437 primary tricompartmental prostheses were compared with 10,624 primary medial or lateral unicompartmental prostheses. *Acta Orthopaedica Scandinavica* 1999;70-2:170-5.
178. **Willis-Owen CA, Brust K, Alsop H, Miraldo M, Cobb JP.** Unicondylar knee arthroplasty in the UK National Health Service: an analysis of candidacy, outcome and cost efficacy. *Knee* 2009;16-6:473-8.

179. **Goodfellow J, O'Connor J.** The mechanics of the knee and prosthesis design. *J Bone Joint Surg Br* 1978;60-B-3:358-69.

180. **Goodfellow JW, Kershaw CJ, Benson MK, O'Connor JJ.** The Oxford Knee for unicompartmental osteoarthritis. The first 103 cases. *J Bone Joint Surg Br* 1988;70-5:692-701.

181. **Price AJ, Svard U.** A second decade lifetable survival analysis of the Oxford unicompartmental knee arthroplasty. *Clinical Orthopaedics and Related Research* 2011;469-1:174-9.

182. **Lisowski LA, van den Bekerom MP, Pilot P, van Dijk CN, Lisowski AE.** Oxford Phase 3 unicompartmental knee arthroplasty: medium-term results of a minimally invasive surgical procedure. *Knee Surgery, Sports Traumatology, Arthroscopy* 2011;19-2:277-84.

183. **Pandit H, Jenkins C, Gill HS, Barker K, Dodd CA, Murray DW.** Minimally invasive Oxford phase 3 unicompartmental knee replacement: results of 1000 cases. *J Bone Joint Surg Br* 2011;93-2:198-204.

184. **Goodfellow JW, O'Connor JJ, Murray DW.** A critique of revision rate as an outcome measure: re-interpretation of knee joint registry data. *J Bone Joint Surg Br* 2010;92-12:1628-31.

185. **Rees JL, Price AJ, Beard DJ, Dodd CA, Murray DW.** Minimally invasive Oxford unicompartmental knee arthroplasty: functional results at 1 year and the effect of surgical inexperience. *Knee* 2004;11-5:363-7.

186. **Hamilton WG, Ammeen D, Engh CA, Jr., Engh GA.** Learning curve with minimally invasive unicompartmental knee arthroplasty. *Journal of Arthroplasty* 2010;25-5:735-40.

187. **Luscombe KL, Lim J, Jones PW, White SH.** Minimally invasive Oxford medial unicompartmental knee arthroplasty. A note of caution! *International Orthopaedics* 2007;31-3:321-4.

188. **Vardi G, Strover AE.** Early complications of unicompartmental knee replacement: the Droitwich experience. *Knee* 2004;11-5:389-94.

189. **Dervin GF, Carruthers C, Feibel RJ, Giachino AA, Kim PR, Thurston PR.** Initial experience with the oxford unicompartmental knee arthroplasty. *Journal of Arthroplasty* 2011;26-2:192-7.

190. **Lindstrand A, Stenstrom A, Ryd L, Toksvig-Larsen S.** The introduction period of unicompartmental knee arthroplasty is critical: a clinical, clinical multicentered, and radiostereometric study of 251 Duracon unicompartmental knee arthroplasties. *Journal of Arthroplasty* 2000;15-5:608-16.

191. **Epinette JA, Brunschweiler B, Mertl P, Mole D, Cazenave A.** Unicompartmental knee arthroplasty modes of failure: wear is not the main reason for failure: a multicentre study of 418 failed knees. *Orthopaedics & Traumatology, Surgery & Research* 2012;98-6 Suppl:S124-30.

192. **Robertsson O, Knutson K, Lewold S, Lidgren L.** The routine of surgical management reduces failure after unicompartmental knee arthroplasty. *J Bone Joint Surg Br* 2001;83-1:45-9.

193. Biomet Skills Academy: Instructional Courses for Medical Professionals.

194. **Cobb J, Henckel J, Gomes P, Harris S, Jakopec M, Rodriguez F, Barrett A, Davies B.** Hands-on robotic unicompartmental knee replacement: a prospective, randomised controlled study of the acrobot system. *J Bone Joint Surg Br* 2006;88-2:188-97.

195. **Citak M, Suero EM, Dunbar NJ, Branch SH, Conditt MA, Banks SA, Pearle AD.** Unicompartmental knee arthroplasty: is robotic technology more accurate than conventional technique? *Knee* 2013;20-4:268-71.

196. **Dunbar NJ, Roche MW, Park BH, Branch SH, Conditt MA, Banks SA.** Accuracy of dynamic tactile-guided unicompartmental knee arthroplasty. *Journal of Arthroplasty* 2012;27-5:803-8 e1.

197. **Karia M, Masjedi M, Andrews B, Jaffry Z, Cobb J.** Robotic assistance enables inexperienced surgeons to perform unicompartmental knee arthroplasties on dry bone models with accuracy superior to conventional methods. *Advances in Orthopedics* 2013;2013:481039.

198. **Jenny JY, Ciobanu E, Boeri C.** The rationale for navigated minimally invasive unicompartmental knee replacement. *Clinical Orthopaedics and Related Research* 2007;463:58-62.

199. **Seon JK, Song EK, Park SJ, Yoon TR, Lee KB, Jung ST.** Comparison of minimally invasive unicompartmental knee arthroplasty with or without a navigation system. *Journal of Arthroplasty* 2009;24-3:351-7.

200. **Weber P, Crispin A, Schmidutz F, Utzschneider S, Pietschmann MF, Jansson V, Muller PE.** Improved accuracy in computer-assisted unicondylar knee arthroplasty: a meta-analysis. *Knee Surgery, Sports Traumatology, Arthroscopy* 2013;21-11:2453-61.

201. **Bauwens K, Matthes G, Wich M, Gebhard F, Hanson B, Ekkernkamp A, Stengel D.** Navigated total knee replacement. A meta-analysis. *J Bone Joint Surg Am* 2007;89-2:261-9.

202. **Gulhane S, Holloway I, Bartlett M.** A vascular complication in computer navigated total knee arthroplasty. *Indian Journal of Orthopaedics* 2013;47-1:98-100.

203. **Novak EJ, Silverstein MD, Bozic KJ.** The cost-effectiveness of computer-assisted navigation in total knee arthroplasty. *J Bone Joint Surg Am* 2007;89-11:2389-97.
204. **Beldame J, Boisrenoult P, Beaufils P.** Pin track induced fractures around computer-assisted TKA. *Orthopaedics & Traumatology, Surgery & Research* 2010;96-3:249-55.
205. **Bonutti PM, Dethmers D, Ulrich SD, Seyler TM, Mont MA.** Computer navigation-assisted versus minimally invasive TKA: benefits and drawbacks. *Clinical Orthopaedics and Related Research* 2008;466-11:2756-62.
206. **Hoke D, Jafari SM, Orozco F, Ong A.** Tibial shaft stress fractures resulting from placement of navigation tracker pins. *Journal of Arthroplasty* 2011;26-3:504 e5-8.
207. **Sonnadara R, McQueen S, Mironova P, Safir O, Nousiainen M, Ferguson P, Alman B, Kraemer W, Reznick R.** Reflections on current methods for evaluating skills during joint replacement surgery: a scoping review. *bone & Joint Journal* 2013;95-B-11:1445-9.
208. **Doyle JD, Webber EM, Sidhu RS.** A universal global rating scale for the evaluation of technical skills in the operating room. *Am J Surg* 2007;193-5:551-5; discussion 5.
209. **Eubanks TR, Clements RH, Pohl D, Williams N, Schaad DC, Horgan S, Pellegrini C.** An objective scoring system for laparoscopic cholecystectomy. *J Am Coll Surg* 1999;189-6:566-74.
210. **Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D, Fried GM.** A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 2005;190-1:107-13.
211. **Vassiliou MC, Kaneva PA, Poulose BK, Dunkin BJ, Marks JM, Sadik R, Sroka G, Anvari M, Thaler K, Adrales GL, Hazey JW, Lightdale JR, Velanovich V, Swanstrom LL, Mellinger JD, Fried GM.** Global Assessment of Gastrointestinal Endoscopic Skills (GAGES): a valid measurement tool for technical skills in flexible endoscopy. *Surg Endosc* 2010;24-8:1834-41.
212. **Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A.** Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Annals of Surgery* 2008;247-2:372-9.
213. **Dath D, Regehr G, Birch D, Schlachta C, Poulin E, Mamazza J, Reznick R, MacRae HM.** Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. *Surgical Endoscopy* 2004;18-12:1800-4.
214. **Sarker SK, Albrani T, Zaman A, Patel B.** Procedural performance in gastrointestinal endoscopy: an assessment and self-appraisal tool. *American Journal of Surgery* 2008;196-3:450-5.

215. **Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B.** Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. *British Journal of Obstetrics and Gynaecology* 2008;115-7:908-16.

216. **Sarker SK, Kumar I, Delaney C.** Assessing operative performance in advanced laparoscopic colorectal surgery. *World Journal of Surgery* 2010;34-7:1594-603.

217. **Pitts D, Rowley DI, Sher JL.** Assessment of performance in orthopaedic training. *J Bone Joint Surg Br* 2005;87-9:1187-91.

218. **Dalkey N.** Experimental Study of Group Opinion - Delphi Method. *Futures* 1969;1-5:408-26.

219. **Pill J.** Delphi Method - Substance, Context - Critique an an Annotated Bibliography. *Socio-Economic Planning Sciences* 1971;5-1:57-71.

220. **Fink A, Kosecoff J, Chassin M, Brook RH.** Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;74-9:979-83.

221. **Rosengart MR, Nathens AB, Schiff MA.** The identification of criteria to evaluate prehospital trauma care using the Delphi technique. *Journal of Trauma* 2007;62-3:708-13.

222. **Mullen PM.** Delphi: myths and reality. *Journal of Health Organization and Management* 2003;17-1:37-52.

223. **Graham B, Regehr G, Wright JG.** Delphi as a method to establish consensus for diagnostic criteria. *Journal of clinical epidemiology* 2003;56-12:1150-6.

224. **Pijls BG, Dekkers OM, Middeldorp S, Valstar ER, van der Heide HJ, Van der Linden-Van der Zwaag HM, Nelissen RG.** AQUILA: assessment of quality in lower limb arthroplasty. An expert Delphi consensus for total knee and total hip arthroplasty. *BMC Musculoskeletal Disorders* 2011;12:173.

225. **Palter VN, Graafland M, Schijven MP, Grantcharov TP.** Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: a Delphi approach. *Surgery* 2012;151-3:391-7.

226. **Zevin B, Levy JS, Satava RM, Grantcharov TP.** A consensus-based framework for design, validation, and implementation of simulation-based training curricula in surgery. *Journal of the American College of Surgeons* 2012;215-4:580-6 e3.

227. **Palter VN, MacRae HM, Grantcharov TP.** Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology. *Am J Surg* 2011;201-2:251-9.

228. **Palter VN, Grantcharov TP.** A prospective study demonstrating the reliability and validity of two procedure-specific evaluation tools to assess operative competence in laparoscopic colorectal surgery. *Surgical Endoscopy* 2012;26-9:2489-503.

229. **Zevin B, Bonrath EM, Aggarwal R, Dedy NJ, Ahmed N, Grantcharov TP.** Development, feasibility, validity, and reliability of a scale for objective assessment of operative performance in laparoscopic gastric bypass surgery. *Journal of the American College of Surgeons* 2013;216-5:955-65 e8; quiz 1029-31, 33.
230. **Moorthy K, Munz Y, Sarker SK, Darzi A.** Objective assessment of technical skills in surgery. *BMJ* 2003;327-7422:1032-7.
231. **Sprague S, Quigley L, Bhandari M.** Survey design in orthopaedic surgery: getting surgeons to respond. *J Bone Joint Surg Am* 2009;91 Suppl 3:27-34.
232. **Stone DH.** Design a questionnaire. *BMJ* 1993;307-6914:1264-6.
233. **Boynton PM, Greenhalgh T.** Selecting, designing, and developing your questionnaire. *BMJ* 2004;328-7451:1312-5.
234. **Rattray J, Jones MC.** Essential elements of questionnaire design and development. *Journal of Clinical Nursing* 2007;16-2:234-43.
235. **Cronbach LJ.** Coefficient Alpha and the Internal Structure of Tests. *Psychometrika* 1951;16-3:297-334.
236. **Nunnally JC.** *Psychometric theory.* 2nd ed ed. New York ; London: McGraw-Hill, 1978:xv, 701 p.
237. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics* 1977;33-1:159-74.
238. **Sugden C, Aggarwal R.** Assessment and feedback in the skills laboratory and operating room. *Surgical Clinics of North America* 2010;90-3:519-33.
239. **Peyre SE, Peyre CG, Hagen JA, Sullivan ME, Lipham JC, Demeester SR, Peters JH, Demeester TR.** Laparoscopic Nissen fundoplication assessment: task analysis as a model for the development of a procedural checklist. *Surgical Endoscopy* 2009;23-6:1227-32.
240. **Bayona S, Akhtar K, Gupte C, Emery RJ, Dodds AL, Bello F.** Assessing Performance in Shoulder Arthroscopy: The Imperial Global Arthroscopy Rating Scale (IGARS). *The Journal of bone and joint surgery. American volume* 2014;96-13:e112.
241. **Sullivan ME, Ortega A, Wasserberg N, Kaufman H, Nyquist J, Clark R.** Assessing the teaching of procedural skills: can cognitive task analysis add to our traditional teaching methods? *American journal of surgery* 2008;195-1:20-3.
242. **Smink DS, Peyre SE, Soybel DI, Tavakkolizadeh A, Vernon AH, Anastakis DJ.** Utilization of a cognitive task analysis for laparoscopic appendectomy to identify differentiated intraoperative teaching objectives. *American journal of surgery* 2012;203-4:540-5.

243. **Cannon-Bowers J, Bowers C, Stout R, Ricci K, Hildabrand A.** Using Cognitive Task Analysis to Develop Simulation-Based Training for Medical Tasks. *Military medicine* 2013;178-10:15-21.

244. **Jacoby J, Matell MS.** Three-Point Likert Scales Are Good Enough. *Journal of Marketing Research* 1971;8-4:495-500.

245. **Burns KE, Duffett M, Kho ME, Meade MO, Adhikari NK, Sinuff T, Cook DJ.** A guide for the design and conduct of self-administered surveys of clinicians. *Canadian Medical Association Journal* 2008;179-3:245-52.

246. **Sierles FS.** How to do research with self-administered surveys. *Academic Psychiatry* 2003;27-2:104-13.

247. **Ewald FC.** The Knee Society total knee arthroplasty roentgenographic evaluation and scoring system. *Clinical Orthopaedics and Related Research* 1989-248:9-12.

248. **Benjamin J.** Component alignment in total knee arthroplasty. *Instr Course Lect* 2006;55:405-12.

249. **Petersen TL, Engh GA.** Radiographic assessment of knee alignment after total knee arthroplasty. *Journal of Arthroplasty* 1988;3-1:67-72.

250. **Bargren JH, Blaha JD, Freeman MA.** Alignment in total knee arthroplasty. Correlated biomechanical and clinical observations. *Clin Orthop Relat Res* 1983-173:178-83.

251. **Ritter MA, Faris PM, Keating EM, Meding JB.** Postoperative alignment of total knee replacement. Its effect on survival. *Clinical Orthopaedics and Related Research* 1994-299:153-6.

252. **Sharkey PF, Hozack WJ, Rothman RH, Shastri S, Jacoby SM.** Insall Award paper. Why are total knee arthroplasties failing today? *Clinical Orthopaedics and Related Research* 2002-404:7-13.

253. **Chau R, Gulati A, Pandit H, Beard DJ, Price AJ, Dodd CA, Gill HS, Murray DW.** Tibial component overhang following unicompartmental knee replacement--does it matter? *Knee* 2009;16-5:310-3.

254. **Gulati A, Chau R, Simpson DJ, Dodd CA, Gill HS, Murray DW.** Influence of component alignment on outcome for unicompartmental knee replacement. *Knee* 2009;16-3:196-9.

255. **Clarius M, Hauck C, Seeger JB, Pritsch M, Merle C, Aldinger PR.** Correlation of positioning and clinical results in Oxford UKA. *Int Orthop* 2010;34-8:1145-51.

256. **Dawson J, Fitzpatrick R, Murray D, Carr A.** Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br* 1998;80-1:63-9.

257. **Weale AE, Halabi OA, Jones PW, White SH.** Perceptions of outcomes after unicompartmental and total knee replacements. *Clinical Orthopaedics and Related Research* 2001-382:143-53.

258. **Bini S, Khatod M, Cafri G, Chen Y, Paxton EW.** Surgeon, implant, and patient variables may explain variability in early revision rates reported for unicompartmental arthroplasty. *J Bone Joint Surg Am* 2013;95-24:2195-202.

259. **Kuipers BM, Kollen BJ, Bots PC, Burger BJ, van Raay JJ, Tulp NJ, Verheyen CC.** Factors associated with reduced early survival in the Oxford phase III medial unicompartment knee replacement. *Knee* 2010;17-1:48-52.

260. **Murray DW, Pandit H, Weston-Simons JS, Jenkins C, Gill HS, Lombardi AV, Dodd CAF, Berend KR.** Does body mass index affect the outcome of unicompartmental knee replacement? *Knee* 2013;20-6:461-5.

261. **Pandit H, Jenkins C, Gill HS, Smith G, Price AJ, Dodd CAF, Murray DW.** Unnecessary contraindications for mobile-bearing unicompartmental knee replacement. *J Bone Joint Surg Br* 2011;93B-5:622-8.

262. **Morgan-Jones R, Haddad FS.** Is this the era of consensus? *Bone & Joint Journal* 2013;95B-11:1441-2.

263. **Shakespeare D, Ledger M, Kinzel V.** The influence of the tibial sagittal cut on component position in the Oxford knee. *Knee* 2005;12-3:169-76.

264. **Shakespeare D, Ledger M, Kinzel V.** Accuracy of implantation of components in the Oxford knee using the minimally invasive approach. *Knee* 2005;12-6:405-9.

265. **Edmondson MC, Isaac D, Wijeratna M, Brink S, Gibb P, Skinner P.** Oxford unicompartmental knee arthroplasty: medial pain and functional outcome in the medium term. *Journal of Orthopaedic Surgery and Research* 2011;6:52.

266. **Pandit H, Murray DW, Dodd CA, Deo S, Waite J, Goodfellow J, Gibbons CL.** Medial tibial plateau fracture and the Oxford unicompartmental knee. *Orthopedics* 2007;30-5 Suppl:28-31.

267. **Van Loon P, de Munnynck B, Bellemans J.** Periprosthetic fracture of the tibial plateau after unicompartmental knee arthroplasty. *Acta Orthopaedica Belgica* 2006;72-3:369-74.

268. **Kim KT, Lee S, Cho KH, Kim KS.** Fracture of the medial femoral condyle after unicompartmental knee arthroplasty. *Journal of Arthroplasty* 2009;24-7:1143 e21-4.

269. **Kendrick BJ, Longino D, Pandit H, Svard U, Gill HS, Dodd CA, Murray DW, Price AJ.** Polyethylene wear in Oxford unicompartmental knee replacement: a retrieval study of 47 bearings. *J Bone Joint Surg Br* 2010;92-3:367-73.

270. **Pegg E, Pandit H, Gill HS, Keys GW, Svard UG, O'Connor JJ, Murray DW.** Examination of ten fractured Oxford unicompartmental knee bearings. *J Bone Joint Surg Br* 2011;93-12:1610-6.

271. **Clarius M, Aldinger PR, Bruckner T, Seeger JB.** Saw cuts in unicompartmental knee arthroplasty: an analysis of Sawbone preparations. *Knee* 2009;16-5:314-6.

272. **Sloper PJ, Hing CB, Donell ST, Glasgow MM.** Intra-operative tibial plateau fracture during unicompartmental knee replacement: a case report. *Knee* 2003;10-4:367-9.

273. **Rudol G, Jackson MP, James SE.** Medial tibial plateau fracture complicating unicompartmental knee arthroplasty. *Journal of Arthroplasty* 2007;22-1:148-50.

274. **Clarius M, Haas D, Aldinger PR, Jaeger S, Jakubowitz E, Seeger JB.** Periprosthetic tibial fractures in unicompartmental knee arthroplasty as a function of extended sagittal saw cuts: an experimental study. *Knee* 2010;17-1:57-60.

275. **Nasca TJ, Day SH, Amis ES, Jr.** The new recommendations on duty hours from the ACGME Task Force. *N Engl J Med* 2010;363-2:e3.

276. **Taffinder N.** Better surgical training in shorter hours. *J R Soc Med* 1999;92-7:329-31.

277. **Grantcharov TP, Reznick RK.** Training tomorrow's surgeons: what are we looking for and how can we achieve it? *ANZ J Surg* 2009;79-3:104-7.

278. **Dumon KR, Traynor O, Broos P, Gruwez JA, Darzi AW, Williams NN.** Surgical education in the new millennium: the European perspective. *Surg Clin North Am* 2004;84-6:1471-91, viii.

279. **Palter VN, Grantcharov TP.** Simulation in surgical education. *Canadian Medical Association Journal* 2010;182-11:1191-6.

280. **Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, Reznick RK.** Randomized controlled trial of virtual reality simulator training: transfer to live patients. *American journal of surgery* 2007;194-2:205-11.

281. **Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM.** Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. *American Journal of Surgery* 2010;199-1:115-20.

282. **Cannon-Bowers JA, Bowers C, Procci K.** Optimizing learning in surgical simulations: guidelines from the science of learning and human performance. *Surgical Clinics of North America* 2010;90-3:583-603.

283. **Yule S, Flin R, Paterson-Brown S, Maran N.** Non-technical skills for surgeons in the operating room: a review of the literature. *Surgery* 2006;139-2:140-9.

284. **Sharma B, Mishra A, Aggarwal R, Grantcharov TP.** Non-technical skills assessment in surgery. *Surg Oncol* 2011;20-3:169-77.

285. **de Leval MR, Carthey J, Wright DJ, Farewell VT, Reason JT.** Human factors and cardiac surgery: a multicenter study. *J Thorac Cardiovasc Surg* 2000;119-4 Pt 1:661-72.

286. **Loveday BP, Oosthuizen GV, Diener BS, Windsor JA.** A randomized trial evaluating a cognitive simulator for laparoscopic appendectomy. *ANZ J Surg* 2010;80-9:588-94.

287. **Schell SR, Flynn TC.** Web-based minimally invasive surgery training: competency assessment in PGY 1-2 surgical residents. *Current Surgery* 2004;61-1:120-4.

288. Ogre3D: http://www.ogre3d.org. (Last accessed March 2014).

289. **Svanaes D, Das A, Alsos OA.** The contextual nature of usability and its relevance to medical informatics. *Studies in Health Technology and Informatics* 2008;136:541-6.

290. **Blyth P, Anderson IA, Stott NS.** Virtual reality simulators in orthopedic surgery: what do the surgeons think? *Journal of Surgical Research* 2006;131-1:133-9; discussion 40-2.

291. **Cohen J.** *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, N.J. ; Hove: Erlbaum Associates, 1988:xxi, 567 p.

292. **Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hamstra SJ.** Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA : Journal of the American Medical Association* 2011;306-9:978-88.

293. **Spencer FC.** Teaching and measuring surgical techniques: the technical evaluation of competence. . *Bull Am Coll Surg* 1978;64:9-12.

294. **Cuschieri A, Francis N, Crosby J, Hanna GB.** What do master surgeons think of surgical competence and revalidation? *Am J Surg* 2001;182-2:110-6.

295. **Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR.** Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database Syst Rev* 2009-1:CD006575.

296. **Hall JC, Ellis C, Hamdorf J.** Surgeons and cognitive processes. *British Journal of Surgery* 2003;90-1:10-6.

297. **Yule S, Paterson-Brown S.** Surgeons' non-technical skills. *Surgical Clinics of North America* 2012;92-1:37-50.

298. **Kahol K, Vankipuram M, Smith ML.** Cognitive simulators for medical education and training. *Journal of Biomedical Informatics* 2009;42-4:593-604.

299. **Subramanian A, Timberlake M, Mittakanti H, Lara M, Brandt ML.** Novel educational approach for medical students: improved retention rates using interactive medical software compared with traditional lecture-based format. *Journal of Surgical Education* 2012;69-4:449-52.

300. Touch Surgery: https://http://www.touchsurgery.com/ (Last accessed October 2014).
301. **Tang B, Hanna GB, Cuschieri A.** Analysis of errors enacted by surgical trainees during skills training courses. *Surgery* 2005;138-1:14-20.
302. **Alvand A, Gill HS, Price AJ, Dodd CA, Murray DW, Rees JL.** Does a mixed training course on the Oxford unicompartmental knee arthroplasty improve non-technical skills of orthopaedic surgeons? *Journal of Orthopaedic Surgery* 2012;20-3:356-60.
303. **Zieky MJ.** So much has changed: how the setting of cutscores has evolved since the 1980s. In: Cizek GJ, Sternberg, R.J., ed. *Setting Performance Standards: Concepts, Methods and Perspectives*, 1 ed. Mahway NJ: Routledge, 2001.
304. **Jordan JA, Gallagher AG, McGuigan J, McClure N.** Virtual reality training leads to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons. *Surg Endosc* 2001;15-10:1080-4.
305. **Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P.** Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg* 2004;91-2:146-50.
306. **Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ.** A systematic review of skills transfer after surgical simulation training. *Ann Surg* 2008;248-2:166-79.
307. **Sabri H, Cowan B, Kapralos B, Porte M, Backstein D, Dubrowskie A.** Serious games for knee replacement surgery procedure education and training. *Innovation and Creativity in Education* 2010;2-2:3483-8.
308. **Van Herzeele I, Aggarwal R, Neequaye S, Darzi A, Vermassen F, Cheshire NJ.** Cognitive training improves clinically relevant outcomes during simulated endovascular procedures. *J Vasc Surg* 2008;48-5:1223-30, 30 e1.
309. **Naik VN, Matsumoto ED, Houston PL, Hamstra SJ, Yeung RY, Mallon JS, Martire TM.** Fiberoptic orotracheal intubation on anesthetized patients: do manipulation skills learned on a simple model transfer into the operating room? *Anesthesiology* 2001;95-2:343-8.
310. **Scott DJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Valentine RJ, Euhus DM, Jeyarajah DR, Thompson WM, Jones DB.** Laparoscopic training on bench models: better and more cost effective than operating room experience? *Journal of the American College of Surgeons* 2000;191-3:272-83.
311. **Meyerson SL, Tong BC, Balderson SS, D'Amico TA, Phillips JD, DeCamp MM, DaRosa DA.** Needs assessment for an errors-based curriculum on thoracoscopic lobectomy. *The Annals of thoracic surgery* 2012;94-2:368-73.

312. **Pugh C, Plachta S, Auyang E, Pryor A, Hungness E.** Outcome measures for surgical simulators: is the focus on technical skills the best approach? *Surgery* 2010;147-5:646-54.
313. **Chandler P, Sweller J.** Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction* 1991;8-4:293-332.
314. **Satava RM.** Identification and reduction of surgical error using simulation. *Minim Invasive Ther Allied Technol* 2005;14-4:257-61.
315. **Arora S, Sevdalis N, Nestel D, Woloshynowych M, Darzi A, Kneebone R.** The impact of stress on surgical performance: a systematic review of the literature. *Surgery* 2010;147-3:318-30, 30 e1-6.
316. **Prabhu A, Smith W, Yurko Y, Acker C, Stefanidis D.** Increased stress levels may explain the incomplete transfer of simulator-acquired skill to the operating room. *Surgery* 2010;147-5:640-5.
317. **Summers AN, Rinehart GC, Simpson D, Redlich PN.** Acquisition of surgical skills: a randomized trial of didactic, videotape, and computer-based training. *Surgery* 1999;126-2:330-6.
318. **Jowett N, LeBlanc V, Xeroulis G, MacRae H, Dubrowski A.** Surgical skill acquisition with self-directed practice using computer-based video training. *American journal of surgery* 2007;193-2:237-42.
319. **Nousiainen M, Brydges R, Backstein D, Dubrowski A.** Comparison of expert instruction and computer-based video training in teaching fundamental surgical skills to medical students. *Surgery* 2008;143-4:539-44.
320. **Krishnan SP, Dawood A, Richards R, Henckel J, Hart AJ.** A review of rapid prototyped surgical guides for patient-specific total knee replacement. *J Bone Joint Surg Br* 2012;94-11:1457-61.
321. **Ast MP, Nam D, Haas SB.** Patient-specific instrumentation for total knee arthroplasty: a review. *Orthopedic Clinics of North America* 2012;43-5:e17-22.
322. **Thienpont E, Bellemans J, Delport H, Van Overschelde P, Stuyts B, Brabants K, Victor J.** Patient-specific instruments: industry's innovation with a surgeon's interest. *Knee Surgery, Sports Traumatology, Arthroscopy* 2013;21-10:2227-33.
323. **Bruni D, Iacono F, Russo A, Zaffagnini S, Marcheggiani Muccioli GM, Bignozzi S, Bragonzoni L, Marcacci M.** Minimally invasive unicompartmental knee replacement: retrospective clinical and radiographic evaluation of 83 patients. *Knee Surgery, Sports Traumatology, Arthroscopy* 2010;18-6:710-7.
324. **Fitz W.** Unicompartmental knee arthroplasty with use of novel patient-specific resurfacing implants and personalized jigs. *J Bone Joint Surg Am* 2009;91 Suppl 1:69-76.

325. **Koeck FX, Beckmann J, Luring C, Rath B, Grifka J, Basad E.** Evaluation of implant position and knee alignment after patient-specific unicompartmental knee arthroplasty. *Knee* 2011;18-5:294-9.

326. **Volpi P, Prospero E, Bait C, Cervellin M, Quaglia A, Redaelli A, Denti M.** High accuracy in knee alignment and implant placement in unicompartmental medial knee replacement when using patient-specific instrumentation. *Knee Surgery, Sports Traumatology, Arthroscopy* 2013.

327. **Jaffry Z, Masjedi M, Clarke S, Harris S, Karia M, Andrews B, Cobb J.** Unicompartmental knee arthroplasties: Robot vs. patient specific instrumentation. *Knee* 2014;21-2:428-34.

328. **Bubbar VK, Kreder HJ.** The intention-to-treat principle: a primer for the orthopaedic surgeon. *J Bone Joint Surg Am* 2006;88-9:2097-9.

329. **Montori VM, Guyatt GH.** Intention-to-treat principle. *Canadian Medical Association Journal* 2001;165-10:1339-41.

330. **Jung KA, Kim SJ, Lee SC, Hwang SH, Ahn NK.** Accuracy of implantation during computer-assisted minimally invasive Oxford unicompartmental knee arthroplasty: a comparison with a conventional instrumented technique. *Knee* 2010;17-6:387-91.

331. **Spencer BA, Mont MA, McGrath MS, Boyd B, Mitrick MF.** Initial experience with custom-fit total knee replacement: intra-operative events and long-leg coronal alignment. *International Orthopaedics* 2009;33-6:1571-5.

332. **Victor J, Dujardin J, Vandenneucker H, Arnout N, Bellemans J.** Patient-specific Guides Do Not Improve Accuracy in Total Knee Arthroplasty: A Prospective Randomized Controlled Trial. *Clinical Orthopaedics and Related Research* 2014;472-1:263-71.

333. **Ng VY, DeClaire JH, Berend KR, Gulick BC, Lombardi AV, Jr.** Improved accuracy of alignment with patient-specific positioning guides compared with manual instrumentation in TKA. *Clinical Orthopaedics and Related Research* 2012;470-1:99-107.

334. **Noble JW, Jr., Moore CA, Liu N.** The value of patient-matched instrumentation in total knee arthroplasty. *Journal of Arthroplasty* 2012;27-1:153-5.

335. **Nunley RM, Ellison BS, Zhu J, Ruh EL, Howell SM, Barrack RL.** Do patient-specific guides improve coronal alignment in total knee arthroplasty? *Clinical Orthopaedics and Related Research* 2012;470-3:895-902.

336. **Barrack RL, Ruh EL, Williams BM, Ford AD, Foreman K, Nunley RM.** Patient specific cutting blocks are currently of no proven value. *J Bone Joint Surg Br* 2012;94-11 Suppl A:95-9.

337. **Nunley RM, Ellison BS, Ruh EL, Williams BM, Foreman K, Ford AD, Barrack RL.** Are patient-specific cutting blocks cost-effective for total knee arthroplasty? *Clinical Orthopaedics and Related Research* 2012;470-3:889-94.

338. **Chareancholvanich K, Narkbunnam R, Pornrattanamaneewong C.** A prospective randomised controlled study of patient-specific cutting guides compared with conventional instrumentation in total knee replacement. *Bone & Joint Journal* 2013;95-B-3:354-9.