

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/148305>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

**CADAVERIC SIMULATION FOR
POSTGRADUATE SPECIALIST TRAINING OF
TRAUMA & ORTHOPAEDIC SURGEONS**

**Hannah K James
MBCChB(Hons) MRCS MMedEd**

**Thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Clinical Education**

**Warwick Medical School, University of Warwick
March 2020**

For body donors.

I hope the work in this thesis contributes in some way towards honouring your extraordinary gift.

Table of contents

List of tables.....	6
List of Figures.....	8
Acknowledgements	11
Declarations	12
Publications	14
Abstract.....	16
Abbreviations	17
Research Training.....	20
Research Training Courses.....	20
Chapter 1: Introduction and background	22
1.1 Introduction	23
1.2 Thesis Aims and Objectives	50
1.3 Potential implications of this work.....	52
1.4 Conclusion	53
Chapter 2: The emerging role of simulation for training surgeons	54
2.1 Introduction	55
2.2 Motor skill acquisition theory.....	56
2.3 Taxonomy of simulation in surgery.....	65
2.4 Educational theory supporting the use of simulation for training surgeons ..	68
2.5 The case for cadaveric simulation in surgery	74
2.6 Conclusion	81
2.7 Reflections for chapters 1 and 2	82
Chapter 3: Cadaveric simulation for postgraduate surgical training: systematic review of the literature	85
3.1 Introduction	86
3.2 Aim	87
3.3 Methods	87
3.4 Results	91
3.5 Discussion.....	123
3.6 Conclusion.....	126
3.7 Reflections	127

Chapter 4: The current status of simulation provision in UK and Republic of Ireland Trauma & Orthopaedic specialist training programmes: a national survey	128
4.1 Introduction	129
4.2 Methods	130
4.3 Results	132
4.4 Discussion.....	147
4.5 Conclusion.....	150
4.6 Reflections.....	151
Chapter 5: Assessment of skill acquisition and operative competency in Trauma & Orthopaedic training	152
5.1 Introduction	153
5.2 Aim	154
5.3 Methods	155
5.4 Results	157
5.5 Discussion.....	204
5.6 Conclusion.....	210
5.7 Reflections.....	211
6. Development of an objective outcome measure of real-world technical operative success using post-operative radiographs.....	212
6.1 Introduction	213
6.2 Aim	216
6.3 Consensus setting exercise	217
6.4 Discussion.....	224
6.5 Conclusion.....	230
6.6 Reflections.....	231
Chapter 7: CAD:TRAUMA Study: Cadaveric simulation versus standard training for postgraduate trauma and orthopaedic surgical trainees; a multi-centre randomised controlled educational trial.....	232
7.1 Introduction	233
7.2 Aim	233
7.3 Objectives	233
7.4 Methods	234
7.5 Results	251
7.6 Discussion.....	294
7.7 Conclusions	306
7.8 Reflections.....	307

Chapter 8: Cadaveric simulation for orthopaedic surgeons-in-training: how does it influence learning and why? A qualitative study.....	311
8.1 Introduction	312
8.2 Research Questions.....	312
8.3 Paradigms: Ontological and Epistemological perspectives.....	313
8.4 Methodology.....	315
8.5 Findings (results).....	319
8.6 Discussion.....	348
8.7 Conclusions	350
8.8 Reflections.....	352
Chapter 9: Discussion and Conclusions.....	355
9.1 Review of thesis objectives	356
9.2 Summary of new findings and context within existing knowledge.....	356
9.3 Conclusions	364
9.4 Future research	365
10. References.....	366
11. Appendix.....	395
11.1 Chapter 3: Data extraction form	395
11.2 Chapter 4: Dashboard survey questions	398
11.3 Chapter 5: Data extraction form	400
11.4 Chapter 6: Consensus setting survey questions.....	402
11.5 Chapter 7: CAD:TRAUMA radiology manual	417
11.6 Chapter 7: Intra-operative blood loss calculator.....	430
11.7 Chapter 7: Sensitivity Analysis	431
11.8 Chapter 8: Topic guide for interviews.....	433

List of tables

Table 1:	Indicative search strategy.....	p.88
Table 2:	Included studies grouped by surgical specialty.....	p.93
Table 3:	Studies by Kirkpatrick level; OCEBM classification and MERSQI scores (range and mean).....	p.94
Table 4a:	Kirkpatrick Level 1: Studies that subjectively measure the impact of the training intervention by learner opinion.....	p.95-100
Table 4b:	Kirkpatrick Level 2: Studies that objectively measure the impact of the training intervention by learner knowledge.....	p.101-102
Table 4c:	Kirkpatrick Level 3: Studies that objectively measure the impact of the training intervention by change in learner behaviour.....	p.103-111
Table 4d:	Kirkpatrick Level 4: Studies that objectively measure the impact of the training intervention by change in patient outcome.....	p.112
Table 5:	Cadaveric models used in the review studies.....	p.121
Table 6:	Van der Vleuten's Utility Index.....	p.154
Table 7:	Example search strategy.....	p.155
Table 8:	Studies assessing arthroscopic performance.....	p.161-175
Table 9:	Studies assessing open surgical performance.....	p.176-187
Table 10:	Utility evidence of the assessment tools used to evaluate surgical competency in T&O.....	p.188-203
Table 11:	Final items.....	p.222-224
Table 12:	Baseline participant characteristics.....	p.253
Table 13:	Tip-Apex distance and lag screw zone position by study group...p.258	
Table 14:	DHS radiographic secondary outcome measures by group.....	p.259
Table 15:	Number of DHS patients experiencing complications.....	p.263
Table 16:	Length of inpatient stay by group.....	p.264
Table 17:	12-month mortality rate by group.....	p.264
Table 18:	Rate of LLD acceptability by group.....	p.267
Table 19:	Number of procedures with acceptable alignment by group.....	p.269
Table 20:	Number of cases with acceptable offset by group.....	p.271

Table 21:	Number of cases of good and poor quality cementation by study group.....	p.273
Table 22:	Blood transfusion rates by study group.....	p.276
Table 23:	Post-operative complication rate by group allocation.....	p.277
Table 24:	Length of inpatient stay by group.....	p.278
Table 25:	12-month mortality rate by group.....	p.278
Table 26:	Acceptability of medial clear space measurements by group.....	p.282
Table 27:	Acceptability of lateral malleolus displacement by group.....	p.284
Table 28:	Acceptability of medial malleolus displacement by group.....	p.286
Table 29:	Acceptability of tibiofibular clear space by group.....	p.288
Table 30:	Acceptability of talocrural angle by group.....	p.290
Table 31:	Demographic details of study participants.....	p.318
Table 32:	Overview of themes and subthemes.....	p.320
Table 33:	Factors influencing surgeons-in-training access to learning opportunities in the operating theatre.....	p.327
Table 34:	Factors driving learning in CST.....	p.335
Table 35:	The added value of CST.....	p.342

List of Figures

Figure 1:	Fitts and Posner's three stages of motor skill acquisition.....	p.57
Figure 2:	Dreyfus and Dreyfus model.....	p.58
Figure 3:	Curran's simulation enhanced trajectory.....	p.63
Figure 4:	Taxonomy of simulation.....	p.66
Figure 5:	Issenberg's 10 features of effective simulation.....	p.69
Figure 6:	Meller's typology of medical education simulators.....	p.70
Figure 7:	Modified Miller's triangle.....	p.70
Figure 8:	Adapted Kirkpatrick's hierarchy.....	p.71
Figure 9:	Miller's putative relationship between fidelity, cost and training effectiveness.....	p.78
Figure 10:	Alessi modifications to the fidelity-transfer correlations.....	p.79
Figure 11:	PRISMA flow-chart of included studies.....	p.89
Figure 12:	Geographical boundaries of the UK and RoI T&O training programmes.....	p.134
Figure 13:	Map showing resources for simulation by programme.....	p.136
Figure 14:	Map of funding sources for in-programme simulation provision.....	p.138
Figure 15:	Barriers to delivering simulation training within programmes.....	p.139
Figure 16:	Map of formal provision of simulation within the teaching timetable.....	p.141
Figure 17:	Map of simulation provision at the mid-stage of specialist training.....	p.143
Figure 18:	Map of simulation provision at the late stage of specialist training.....	p.144
Figure 19:	Map of activity measuring the impact of in-programme simulation provision.....	p.146
Figure 20:	PRISMA diagram.....	p.158
Figure 21:	Schematic overview of patient post-operative recovery timeline and the factors that can influence outcome from surgery.....	p.214
Figure 22:	Venn diagram showing the required qualities of the radiological measurement parameters.....	p.216
Figure 23:	Overview of the consensus exercise.....	p.221

Figure 24:	Schematic overview of the study timeline.....	p.237
Figure 25:	Schematic overview of course set-up in the surgical training centre.....	p.247
Figure 26:	Course timetable – Day 1: left sided procedures.....	p.248
Figure 27:	Course timetable – Day 2: right sided procedures.....	p.249
Figure 28:	The West Midlands Surgical Training Centre.....	p.250
Figure 29:	A cadaveric DHS being performed during the course.....	p.250
Figure 30:	A cadaveric ankle fracture fixation being performed during the course (two DHS simulations in background).....	p.250
Figure 31:	CONSORT diagram of participant flow.....	p.252
Figure 32:	Diagram of DHS in-situ showing Tip-Apex Distance and femoral head position.....	p.255
Figure 33:	Modified Cleveland’s femoral head zone.....	p.256
Figure 34:	Tip-Apex Distance by study group.....	p.257
Figure 35:	Procedure time by study group.....	p.260
Figure 36:	Intra-operative radiation dose to patient.....	p.262
Figure 37:	Summary diagram of hemiarthroplasty measurements.....	p.265
Figure 38:	Leg length discrepancy by group.....	p.266
Figure 39:	Femoral stem alignment on the post-operative radiograph in the AP view.....	p.268
Figure 40:	Difference in femoral offset between native and surgical hips.....	p.270
Figure 41:	Frequency of Barrack cement mantle quality gradings by study group.....	p.272
Figure 42:	Hemiarthroplasty procedure time by study group allocation.....	p.274
Figure 43:	Volume of intra-operative blood loss by group allocation.....	p.275
Figure 44:	Schematic overview of ankle fracture radiographic measurements.....	p.280
Figure 45:	Medial clear space following ankle fracture fixation by group.....	p.281
Figure 46:	Lateral malleolar displacement over time by study group.....	p.283
Figure 47:	Medial malleolar displacement over time by study group.....	p.285
Figure 48:	Tibiofibular clear space by study group.....	p.287
Figure 49:	Talocrural angle by study group.....	p.289
Figure 50:	Procedure study group.....	p.291

Figure 51: Radiation dose to patient by group over time.....p.293

Acknowledgements

I would like to thank my supervisors Professor Damian Griffin, Mr Giles Pattison and Dr Joanne Fisher for their enduring support, guidance and wisdom during my research degree.

I am also very grateful to the following people for their help and collaboration;

- Local principle investigators and research facilitators who supported me in collecting clinical data from participating hospital sites; Ms Sonia Kandola, Mr Neil Ashwood, Mr Divya Prakash, Ms Connie Rowlands, Ms Jo Davies, Mr Sohail Quraishi, Mr Bobin Varghese, Ms Deepa Bose, Mr Ben Ollivere
- Mrs Anna Chapman for being second reviewer for the two systematic reviews in this thesis
- Mrs Yessica Diez-Davies, who enhanced my diagrams and illustrations using her skills as professional graphic designer
- Ms Samantha Johnson, academic support librarian, for being so helpful and always having the answer
- Dr Nick Parsons, senior medical statistician, for advising me on the analysis

Finally I would like to thank my husband Dan for his endless support, and my three daughters Ottie, Effie and Trixie who were born during the creation of this thesis.

“This is Mummy’s book of spells” said Ottie, age 5. A great description for a PhD thesis!

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree, apart from subsections 2.5.1 'The history of training surgeons using cadavers' and 2.5.2 'Legal position regarding the use of cadavers for surgical training' (chapter 2, p74-76 only) which are partly based on an assignment towards the award of the Masters in Medical Education degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Chapter 2: Mrs Yessica Diez-Davies (YDD), graphic designer, redrew my word processor diagrams using professional graphic design software

Chapter 3: Mrs A W Chapman (AWC), consultant orthopaedic surgeon, provided independent data extraction as second reviewer and independent assessment of references for inclusion. YDD improved the diagrams.

Chapter 4: Mr Rob Gregory (RG), consultant orthopaedic surgeon and Specialist Advisory Committee chair, designed and administered the Performance and Opportunity Dashboard Survey. YDD drew the maps using professional graphical design software based on my hand-drawings.

Chapter 5: AWC provided independent assessment of references for inclusion

Chapter 6: Parts of the ankle core outcome set was adapted with permission from the AIR Trial⁴²⁹. YDD improved the diagrams.

Chapter 7: The cadaveric simulation training course was delivered by a volunteer consultant faculty. My supervisors (GTRP and DRG) helped organise, provision and deliver the course. Dr Nick Parsons, senior medical statistician, advised me on the statistical analysis and checked it was conducted correctly. Mr John Skliros, graphic designer, designed the logo for the trial.

Chapter 8: I used a professional transcription service to transcribe the interview recordings for analysis

Funding:

The original research presented in this thesis was funded by;

- Versus Arthritis Educational Research Fellowship Award (Grant number 20485, £149,998 – for salary)
- Health Education West Midlands Faculty Development Grant (£30,500 – for consumables)
- Chancellor’s Scholarship, University of Warwick (£24,200 – for PhD registration fees)
- CVC Foundation project grant (£68,200 – for consumables and dissemination)
- Honorary Royal College of Surgeons Research Fellowship (research methods training and support)

Publications

Parts of this thesis have been published, or submitted for publication, by the author:

Peer reviewed publications:

James HK, Chapman AW, Pattison GTR, Griffin DR. Assessment of technical skill acquisition and operative competence in Trauma and Orthopedic surgical training: A systematic review. JBJS Reviews. 2020 In Press. doi:10.2106/JBJS.RVW.19.00167

James HK, Fisher JG, Griffin DR, Pattison GTR. If we can't get to theatre, we can't learn to operate. A qualitative study of factors influencing core trainee access to the operating theatre in Trauma & Orthopaedics. Ann R Coll Surg Engl. 2020, In Press.

James HK, Gregory R, Tennent D, Pattison GTR, Fisher JD, Griffin DR. Current provision of simulation in the United Kingdom and Republic of Ireland Trauma & Orthopaedic specialist training: A National Survey. Bone & Joint Open. 2020;1-5:103-114.

James HK, Pattison GTR, Griffin DR, Fisher JD. How does cadaveric simulation influence learning in Orthopedic residents? J Surg Educ. 2020;77(3):671-682.

James HK, Chapman AW, Pattison GTR, Griffin DR, Fisher JD. Systematic review of the current status of cadaveric simulation for surgical training. British Journal of Surgery. 2019;106(13):1726-34.

James HK. Measuring the educational impact of simulation training in Trauma & Orthopaedics. Journal of Trauma & Orthopaedics. 2019;7(3):54-56.

James HK. Use of cadavers to train surgeons: what are the ethical issues? J Med Ethics Epub ahead of print: 2019;0:1-2 doi:10.1136/medethics-2019-105873.

Walker TA, **James HK**. Use of cadavers to train surgeons: what are the ethical issues – body donor perspective. J Med Ethics. 2020;medethics-2019-105998.doi:10.1136/medethics-2019-105998

Submitted for publication:

James HK, Pattison GTR, Fisher JD, Griffin DR. Cadaveric simulation vs standard training for postgraduate trauma & orthopaedic surgical trainees: protocol for the CAD:TRAUMA study multi-centre randomised controlled education trial. Submitted to BMJ Open.

Letters:

James HK, Chapman AW, Pattison GTR, Griffin DR, Fisher JD. Author response to: Comment on: Systematic review of the current status of cadaveric simulation for surgical training. British Journal of Surgery. 2020;107(4):468.

James HK, Pattison GTR. Author response to: Comment on: Systematic review of the current status of cadaveric simulation for surgical training. British Journal of Surgery. 2020;10.1002/bjs.11629.doi:10.1002/bjs.11629.

Conference podium presentations:

James HK. Current provision of simulation in United Kingdom and Republic of Ireland T&O Specialist Training. Keynote lecture, British Orthopaedic Association Annual Congress, Liverpool, September 2019.

James HK. Cadaveric simulation for training Trauma & Orthopaedic Surgeons. Postgraduate Research Symposium, University of Warwick, April 2019.

Abstract

Surgical training is threatened by reduced working hours and a health service that is struggling to meet patient demand. There is a need to maximise training efficiency to ensure the calibre of the consultant surgical workforce is maintained. One solution is using simulation as an adjunct to the traditional master-apprentice training model. Junior surgeons-in-training can be rapidly upskilled in the simulation laboratory, moving the early part of the learning curve away from patients. Cadaveric simulation uses deceased human bodies to teach operations. A systematic review showed that there is an abundance of low quality evidence that cadaveric simulation induces short-term behavioural change when assessed by objective measures. There is a lack of evidence of skill retention longitudinally, of transfer to live theatre and of patient benefit. A national survey of simulation provision in Trauma & Orthopaedic training programmes showed widespread but inconsistent provision of cadaveric simulation, with a complex funding landscape and reliance on industry sponsorship. A systematic review of technical skills assessment tools used in Trauma & Orthopaedics showed the utility evidence for these is weak, and there is inadequate evidence to support the continued use of the procedure based assessment as the gold-standard in high stakes assessment of competency. A consensus exercise developed a core outcome set of clinically relevant radiographic measurements to assess technical operative skills. A multicentre randomised controlled trial comparing the performance of cadaveric vs standard-trained junior surgeons-in-training showed that there were some significant improvements in implant position, acute complication rate and blood transfusion requirement across three common trauma operations. There could be substantial cost savings for the health service if all junior surgeons-in-training received cadaveric simulation. A qualitative study showed that it provides an optimal deliberate practice environment to learn a wide range of technical and non-technical skills in a complete training package.

Word count 300/300

Abbreviations

ACLR	Anterior Cruciate Ligament Reconstruction
ANP	Advanced Nurse Practitioner
AO	Arbeitsgemeinschaft für Osteosynthesfragen
AP	Anteroposterior
ARCP	Annual Review of Competence Progression
ASiT	Association for Surgeons in Training
ASSET	Arthroscopic Surgical Skill Evaluation Tool
BAKSSS	Basic Arthroscopic Knee Skills Scoring System
BEME	Best Evidence Medical Education
BMA	British Medical Association
BOA	British Orthopaedic Association
BOTA	British Orthopaedic Trainee Association
BSREC	Biomedical and Scientific Research Ethics Committee
CAG	Confidentiality Advisory Group
CBI	Competency Based Instruction
CCT	Certificate of completion of training
CCST	Certificate of completion of specialist training
CONSORT	Consolidated Standards of Reporting Trials
COR	Cut out rate
CT	Core Training
CTI	Chest Tube Insertion
CTB	Cadaveric Temporal Bone
CTR	Carpal Tunnel Release
CST	Cadaveric Simulation Training
DCP	Dynamic Compression Plate
DHS	Dynamic Hip Screw
DRFF	Distal Radius Fracture Fixation
EEA	European Economic Area
ENT	Ear, Nose and Throat Surgery
EO	Expert Opinion
EWTD	European Working Time Directive
ETI	Endotracheal Tube Insertion
Ex-Fix	External Fixator
FFC	Fresh Frozen Cadaver
FORS	Fundamentals of Orthopaedic Surgery
FPA	Final Product Analysis
FRCS	Fellowship of the Royal College of Surgeons
FY	Foundation Year Doctor
GMC	General Medical Council
GOALS	Global Operative Assessment of Laparoscopic Skills
GOSLE	Generic Operative Supervised Learning Event
GRRSA	Global Rating Scale for Shoulder Arthroscopy
GRS	Global Rating Scale
GSS	Global Summary Scale
HaN	Hospital at Night
HEE	Health Education England

HMA	Hand Motion Analysis
HRA	Health Research Authority
HTA	Human Tissue Act
IGARS	Imperial Global Arthroscopy Rating Scale
IGI	Injury Grading Index
IM	Intramedullary
IST	Improving Surgical Training
ISRCTN	International Standard Randomised Controlled Trial Number
IT	Information Technology
JCST	Joint Committee on Surgical Training
LC-CUSUM	Cumulative Summation Test for Learning Curve
LLD	Leg Length Discrepancy
LMD	Lateral Malleolar Displacement
LoE	Level of Evidence
LR	Labral Repair
LTFT	Less than full time
MCS	Medial Clear Space
MEE	Medical Education England
MeSH	Medical Subject Heading
MERSQI	Medical Education Research Quality Instrument
mGym ²	Milligray per metre squared
MMC	Modernising Medical Careers
MMD	Medial Malleolar Displacement
MTAS	Medical Training Application Service
NHFD	National Hip Fracture Database
NHS	National Health Service
NJR	National Joint Registry
NOF	Neck of Femur
NOTSS	Non-technical skills for surgeons
NS	Not specified
NTN	National Training Number
OAAS	Objective Assessment of Arthroscopic Skills
OCEBM	Oxford Centre for Evidence of Based Medicine
O&G	Obstetrics and Gynaecology
OPAT	Objective Practical Assessment Tool
ORIF	Open Reduction Internal Fixation
OSATS	Objective Structured Assessment of Technical Skills in Surgery
PACS	Picture Archiving and Communication System
PBA	Procedure Based Assessment
PCT	Primary Care Trust
PGY	Postgraduate Year
PI	Principle Investigator
PMETB	Postgraduate Medical Education and Training Board
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PROMS	Patient Reported Outcome Measures
PROSPERO	International Prospective Register of Systematic Reviews
PT	Procedure Time
RCR	Rotator Cuff Repair
RCT	Randomised Controlled Trial

RoI	Republic of Ireland
SAC	Specialist Advisory Committee
SAS	Staff and Associate Specialist
SBE	Simulation Based Education
SD	Standard Deviation
SDM	Simulator Derived Metrics
SHO	Senior House Officer
SoTR	Shape of Training Review
SpR	Specialist Registrar
ST	Specialist Trainee
TAD	Tip Apex-Distance
TBC	Task Based Checklist
TBW	Tension Band Wire
TCA	Talo-Crural Angle
TEP-IHR	Total extra-peritoneal inguinal hernia repair
TFCS	Tibiofibular Clear Space
TFR	Trigger Finger Release
THA	Total Hip Arthroplasty
T&O	Trauma and Orthopaedic Surgery
TPD	Training Programme Director
TS	Talar Subluxation
TSC	Task Specific Checklist
UHCW	University Hospital Coventry & Warwickshire
UK	United Kingdom
UKSTSG	UK Shape of Training Steering Group
US(A)	United States (of America)
VC	Venous Cutdown
VR	Virtual Reality
WHO	World Health Organisation
WMSTC	West Midlands Surgical Training Centre
3D	3 dimensional

Research Training

Research Training Courses

December 2012	Qualitative and Comparative Research Methods, University of Warwick
November 2013	Clinical Research Methods for Surgeons, Royal College of Surgeons of England Academic Networking, WMS Doctoral Research Training Programme
April 2014	Practical Project Management, WMS Doctoral Research Training Programme
May 2014	Mixed Methods for Health Research, Warwick Medical School Effective Literature Searching, Warwick Medical School
June 2014	Academic Writing 1 – thesis structure, WMS Doctoral Research Training Programme Academic Writing 2 – defending your argument, WMS Doctoral Research Training Programme
June 2018	NVivo Qualitative Software Training, University of Warwick
July 2018	Good Clinical Practice Training, University of Warwick
October 2018	Reference Management using Endnote Desktop Version, University of Warwick
October 2019	IBM SPSS Statistics Training, University of Warwick

Conferences Attended

September 2013	British Orthopaedic Association Annual Congress, Birmingham
March 2014	Arthritis Research UK Annual Fellows Conference, Loughborough
April 2014	International Conference on Surgical Education and Training (ICOSET), Harrogate
April 2017	Arthritis Research UK Annual Fellows Conference, Loughborough
September 2018	British Orthopaedic Association Annual Congress, Birmingham
March 2019	International Conference on Surgical Education and Training (ICOSET), Edinburgh
May 2019	WMS Postgraduate Research Symposium, University of Warwick
September 2019	British Orthopaedic Association Annual Congress, Birmingham
March 2020	Versus Arthritis Annual Fellows Conference, Loughborough

Chapter 1: Introduction and background

In this first chapter I will describe the recent history of surgical training reform in the National Health Service. I will discuss the current challenges facing surgical training delivery and will describe their impact in Trauma & Orthopaedic surgery. Having set the scene for the research, I will state the aims and objectives of this thesis.

Declarations

None

1.1 Introduction

1.1.1 The recent history of surgical training reform

In order to understand how the current surgical training climate evolved, a brief description of the recent history of surgical training in the National Health Service (NHS) is necessary.

1950-1980s: The Halstedian Era

Surgical training in the young, post-war NHS of the 1950s was very different to what we know today^{1 2}. Surgical procedures were performed by generalist surgeons, and the range of procedures was extremely limited by modern standards². There was minimal scope for specialisation and this only occurred after a surgeon had completed training and was practicing independently at consultant level².

Operations were performed using open incisions and there was no such thing as arthroscopic, laparoscopic or minimally invasive surgery. Post-operative recovery of the patient was encouraged by prolonged bedrest rather than active rehabilitation. Patients were not invited to ask questions of their surgeons nor actively participate in decisions about their care, owing to the paternalistic medical culture of the time³.

Hospital inpatient stays were longer, caseloads were lower and surgeons-in-training worked more hours but at a significantly lower intensity, with overall responsibility for fewer patients, whose care they were involved in for the duration of the inpatient stay³.

Surgeons were exposed to more than 30,000 hours of operating time during the course of typical training⁴, and entry and exit points to training were ill-defined. There was no curriculum, no formal assessment process to signpost progression through training, and surgeons-in-training progressed through the ranks when their

superiors deemed them ‘ready’ – a subjective and nepotistic process –with no mechanism for identifying and remediating failing trainees^{5 6}.

i. The master-apprentice model of training

Surgeons-in-training were trained according to the master-apprentice model first described by Sir William Halstead in 1889^{7 8}. Central to this model is a process of ‘graded responsibility’^{7 9}. The apprentice surgeon would observe the master at work and progress from observation, to performing part of a procedure under supervision, to entire procedures under supervision, and finally to independent practice⁹. Success using this model relied on three key factors described by Walter¹⁰; 1) large case numbers with multiple, reproducible opportunities for repetition of skills, 2) long hours of work to ensure adequate exposure to these necessary volumes and 3) consistent, skilled consultant mentors.

Hargreaves¹¹ further divides the master-apprentice model into two categories; ‘apprenticeship-by-osmosis’ and ‘apprenticeship-by-coaching’. Apprenticeship-by-coaching is considered superior¹¹, and to be successful requires an effective master who coaches the apprentice by role-modelling, carefully supervises progress and provides hands-on experience¹¹. Apprenticeship-by-osmosis is described as being inferior as it is characterised by learning opportunities happening in an ‘unplanned, unsystematic and unsupported’ fashion¹¹.

Apprenticeship-by-osmosis has remained the default training method in surgery despite its idiosyncrasies^{10 12}. This is partly because there has historically been a lack of importance attached to the status of trainer (master)¹³ and trainers have been shown to exhibit behaviours they themselves were exposed to as trainees¹⁴, so this style of training perpetuates. It has only recently begun to be recognised that surgeon-trainers ought to have proper training in education, and the role of trainer professionalised and rewarded^{13 15 16}.

In spite of its relative success in training competent surgeons, (particularly when there is an emphasis on apprenticeship-by-coaching), the master-apprentice model is

increasingly being recognised as unsuitable for modern healthcare training environments^{7 9 17}. This is because all three of the central foundations required for the success of the master-apprentice model (large case numbers, long working hours and consistent mentors) are in jeopardy in the modern NHS for reasons I will describe below.

ii. Early postgraduate training regulation efforts

Problems with the unstructured nature of surgical training were recognised early and efforts to introduce regulation were beginning to be made as early as 1950. The Medical Act of 1950^{18 19} made the requirement of the ‘house year’, the first year of supervised clinical practice after graduation, mandatory. The Medical Act of 1956 charged the newly convened General Medical Council (GMC) Education Committee with statutory responsibility for regulating postgraduate training in the UK, and by the 1960’s there was pan-collegiate recognition of all non-consultant posts as ‘training grades’, with the requirement for NHS hospitals to support postgraduate education¹⁹. By the early 1970’s there was an increasing recognition of the need for co-ordination of postgraduate training, which was previously highly heterogeneous and hospital dependent¹⁹. The creation of Postgraduate Training Committees affiliated with a regional university were created to better co-ordinate training¹⁹.

1990s: The Calman Reforms

Formal, UK-wide recognition of the lack of structure and assessment in surgical training (and postgraduate medical training more widely) came in the form of a 1993 parliamentary working group report on specialist medical training entitled ‘Hospital Doctors: Training for the Future’²⁰. The report led to the introduction of the eponymous Calman reforms by the then Chief Medical Officer of England Sir Kenneth Calman (1991-1998). The main aims of these reforms were to bring the UK in line with EU law^{4 19} and to standardise UK based postgraduate surgical (and medical) training with that of the rest of Europe.

The Calman reforms had two main components. Firstly; to restructure postgraduate specialist training by combining the old ‘registrar’ and ‘senior registrar’ grades into the new ‘Specialist Registrar’ grade, and to define entry points to training and to limit training to 7-9 years after primary medical qualification^{19 20}. The Royal Colleges were tasked with providing curricula and specify competency based assessment in progression of training^{19 20}. Secondly the concept of the ‘Certificate of Completion of Specialist Training (CCST)’ was introduced. This was to define the endpoint of training, whereby receipt of the CCST from the General Medical Council on instruction from the Royal College would signify the individual has completed a training programme to a ‘standard compatible with independent practice’^{19 21}. Of note, the Calman reforms impacted the training of doctors only once they had progressed from the Senior House Officer (SHO) grades to Specialist Registrar (SpR).

Mid 1990s - early 2000s: The post-Calman years

i. Senior House Officers – ‘The Lost Tribe’

In the early post-Calmanisation era of the mid 1990’s, concern remained about the lack of structure and progression points for the many doctors in the early stages of their postgraduate training - the ubiquitous ‘Senior House Officer’ or ‘SHO’ doctor^{19 22}.

A consultation paper on medical training reform written in 2002 by Liam Donaldson, Chief Medical Officer, entitled ‘Unfinished Business’ highlighted the key issues²³. Many SHO’s were employed in short term posts, without formally recognised training value. A lack of clearly defined exit points meant many were ‘trapped’ at this level for up to ten years without prospect of progress to Specialist Registrar level, leading to the famous description of the post-Calman SHO’s as the ‘lost tribe’^{22 24}. Job satisfaction and morale were low, and an increasing workload burden and bias towards service provision led to their description as the ‘workhorses’ of the NHS^{22 23}. Five key recommendations were made for the reform of SHO training; that it be programme-based, time limited, broad-based, flexible and individually tailored^{23 24}. The report suggested the introduction of a two year ‘foundation

programme' immediately following graduation from medical school, leading into entry to eight or more broad-based basic specialist training programmes, before competitive entry into specialty Specialist Registrar schemes which, assuming satisfactory progress, would lead to the award of a CCST²³.

ii. Non-training grades and 'Professional cul-de-sacs'

Following the proposals set out in 'Unfinished Business', legislation was tabled by ministers that would lead to the creation of the Postgraduate Medical Education and Training Board (widely known as PMETB), a non-governmental public body responsible for the postgraduate training of doctors²⁵.

Shortly after 'Unfinished Business', a second consultation paper on medical training reform was published in 2003 entitled 'Choice and Opportunity'²⁶. This addressed mounting concerns about difficulties experienced by doctors in stand-alone non-training posts and those outside the formal training system without consultant or GP status – known collectively as Staff and Associate Specialist doctors (SAS doctors)²⁵. These difficulties included diverse role variation, lack of career structure and opportunities for progression, described as 'professional cul-de-sacs'^{24 25}, and a recognised stigma associated with non-training grade positions²⁴. The Choice and Opportunity report detailed recommendations to reform SAS medical careers along the principles of improving access to training, continuous professional development opportunity and recognition of SAS posts as valid career choices^{24 26}.

There were several additional political and legislative factors that influenced the need for training reform in the late 1990s and early 2000s; the so-called 'gathering storm'²⁴.

iii. Changing public expectations of healthcare delivery

The NHS Plan in 2000 detailed a manifesto for a top-heavy model of healthcare delivery that was to be increasingly provided by consultants or fully trained doctors²⁷, rather than the 'workhorse' SHOs²⁴. This increased the need for new

consultants and output of trainees from specialty training schemes, and hence applied pressure to reduce minimum training times²⁴. Medical school numbers expanded by 60% between 1999 and 2005 as part of the drive towards workforce self-sufficiency, and to provide the numbers needed to progress through the postgraduate training schemes²⁴.

iv. Working hours restrictions

Doctors' work hours were capped at 56 hours per week in 1991, as a result of the 'New Deal'²⁸. Minimum rest periods between shifts were enforced²⁹, in recognition of public and professional demand amid the mounting evidence on the negative effects of sleep deprivation and unnatural circadian rhythms. The risk to personal health, cognitive and motor impairment^{30 31}, and the associated risk of clinical errors and injuries³² meant there was potential for serious patient harm as a result of doctor fatigue³³.

The working hours of doctors changed further with the introduction of the European Working Time Directive (EWTD). The EWTD became law in 1993, but doctors, and several other professions, were initially excluded³². In 2004 this exclusion was removed as part of a planned, phased inclusion for formerly exempt workers by the European Commission and Parliament³². In 2004 doctors working hours were limited to 58 hours a week and then in 2009, to 48 hours a week. The EWTD placed additional restrictions on compulsory rest and rota design that were more stringent than The New Deal, including the requirement for 11 hours continuous rest in each 24 hour period³².

Concerns began to be raised about how training could be delivered in greatly reduced hours, and studies during the transition period showed that time spent by trainees in the operating theatre and outpatient clinic – the two principle surgical training environments - was significantly reduced in a EWTD compliant rota system compared to a traditional on-call rota^{29 34}.

v. NHS deficits

2004-5 brought unprecedented financial strain to the NHS in the form of widespread deficits, leading to pressure on all parts of the organisation to reduce spending²⁴. The creation of National Training Levies in 1996 had meant that postgraduate education budgets had been safely ring-fenced from spending cuts, but this protection was removed by the Department of Health in the midst of the austerity measures²⁴, and many Strategic Health Authorities acted on their new freedom to reduce commitment to funding training and education from 2006^{24 35}.

vi. Structural re-organisation within the NHS

Major changes in the funding structure of the NHS occurred in the early 2000s. This principally involved significant reductions in the number of Strategic Health Authorities (SHA's) and Primary Care Trusts (PCT's), which had a temporary destabilising effect on NHS organisations and further distracted from the ongoing issues with postgraduate training²⁴.

vii. The creation of the Postgraduate Medical Education and Training Board

PMETB was created in 2003, and began work in 2005²⁴ as an independent statutory body with overall responsibility for regulating postgraduate medical education and training, within a single framework³⁶. This was part of a wider effort to make professional regulation more centralised and accountable, as a result of high-profile 'self-regulatory' failings including the Bristol Royal Infirmary and Alder Hey scandals²⁴. PMETB took over the responsibility of regulating training from the specialist training authorities of the individual royal colleges³⁶.

2003-7 Modernising Medical Careers (MMC)

i. The vision of Modernising Medical Careers

In 2003 the UK's four Health Departments released a joint policy statement entitled 'Modernising Medical Careers'³⁷. This outlined the principles of an ambitious and major reform for postgraduate training, driven by the issues discussed above. The key features of this reform were scrapping the SHO grade, creating the Foundation Programme³⁸ (compulsory two-year broad based training for all medical graduates), and, most significantly, streamlining specialist training.

A UK Strategy Group was formed in late 2003 by Sir Liam Donaldson to deliver the MMC reforms, which led to the 2004 publication of 'MMC: The next steps – The future shape of Foundation, Specialist and General Practice Training Programmes'³⁹. The report outlined the most dramatic overhauling of the structure of medical education since the start of the NHS in 1948.

A crucial feature of the MMC reform was the introduction of a single, run-through, training track for postgraduate specialist training and general practice. The idea was to converge the previous basic and higher specialist training phases, with the associated competitive entry transition between the two which was resulting in the 'lost tribe' and 'professional cul-de-sac' problems (described above), into a 'seamless training process would lead directly to the award of CCST (assuming satisfactory progress), without additional competitive entry'³⁹. The main principles of the new system were set out as the 'Seven Pillars of MMC'²⁴, that training be; trainee-centred, competency assessed, service-based, quality assured, flexible, coached and streamlined'.

ii. The Medical Training Application Service debacle

The Foundation Programme was launched in 2005, with relative success⁴⁰. In 2007, the MMC programme was fully deployed and all doctors in training (not foundation years) were required to apply for training positions using the newly formed

centralised recruitment system; the Medical Training Application Service (MTAS). There were serious early concerns about MTAS, including grave flaws in the shortlisting process and software failures including confidential data breaches²⁴.

There was also a significant excess of applicants owing to a failure to restrict applications from the non-European Economic Area (EEA). In the ensuing chaos, many highly qualified and experienced junior doctors received no interview offers, and their future careers were left in limbo. The British Medical Association (BMA) described ‘top juniors flung on the scrapheap after a decade of training’²⁴ and that the leaders of the medical profession were ‘hand-maidens to their own apocalypse’²⁴. There was total ‘collapse in confidence’²⁴ in the recruitment and selection process and there followed a series of high profile resignations, including the head of MMC, Alan Crockard²⁴, and heated public demonstrations by junior doctors⁴¹. The Lancet re-designated the mnemonic MMC to ‘mass-medical culling’⁴¹.

An emergency review group was tasked with making urgent reform mid-way through the recruitment process, the unpopular results of which were unsuccessfully challenged in a High Court judicial review⁴² sought by Remedy UK, a now-defunct junior doctors action group²⁴. Several more high-profile resignations from senior figures involved in MMC followed⁴³, and a series of emergency ministerial statements were issued to address concerns about the viability of MMC and the postgraduate training environment as a whole²⁴.

The events of 2007 are widely considered to have been damaging for junior doctors, the reputation of postgraduate training and highly embarrassing for the government²⁴.

iii. The Tooke Inquiry

As a consequence of the MTAS fiasco, an independent inquiry into the events of 2007 was led by Professor Sir John Tooke, the final report of which was published in 2008⁴⁴. The inquiry found that the 2007 postgraduate training crisis had occurred for several reasons; poor policy objectives, weak governance structure, inadequate project management and absent leadership initiative from both the medical

profession and Department of Health²⁴. Ineffective workforce planning for the new streamlined training structure further compounded the problems²⁴.

The Tooke inquiry made 47 recommendations for reform to postgraduate medical education. It called for a more ‘flexible and broad based approach to training, integrating both training and service into workforce planning’⁴⁴. These recommendations, crucially, included key structural training reforms which directly opposed the two key changes made by MMC; splitting the foundation programme and uncoupling basic and higher specialist training at registrar level. A new training body was also proposed – NHS Medical Education England (NHS: MEE) to act as the ‘policy interface between development and interface on matters related to postgraduate medical education and training’ and as principle budget holder for postgraduate training^{44 45}.

The response to the Tooke report from the Royal Colleges^{46 47}, the BMA, NHS Employers and doctors themselves was overwhelming positive^{24 48}. The Department of Health welcomed the report and declared it a ‘significant step forward’²⁴ in addressing the ongoing issues with postgraduate training following the 2007 crisis.

2009-2018; The post-Modernising Medical Careers, ‘Shape-of-Training’ climate

Postgraduate training and recruitment continued in a precarious state of equilibrium during the years following the crisis of 2007. Many workforce deaneries began offering ‘mixed-economy’ schemes, which had a quota of MMC-style ‘run-through’ specialist training posts within their intake, with the rest of trainees being appointed into the new ‘uncoupled’ style posts. Upon completion of foundation training, trainees would be appointed by competitive application into a ‘core’ surgical training post for two years, with the training years called ‘CT1’ and ‘CT2’ respectively. During CT2, trainees would go through a further process of open, competitive national recruitment into the higher specialist training schemes, beginning at the third specialist training year ‘ST3’. Once appointed into an ST3 post with a National Training Number (NTN), trainees would progress through training for a further 5 years until CCST at the end of ST8 (ST7 in general surgery). This mixed-economy

structure attracted early criticism that uncoupled trainees would be unfairly disadvantaged⁴⁹.

i. Shape of Training Review

The Shape of Training Review ‘SoTR’, led by economist Professor David Greenaway, was undertaken in 2013. The remit was to establish how postgraduate training could be improved to meet the projected needs of patients and service providers over the next 30 years, and to learn from lessons following MMC⁵⁰. The review made 19 recommendations for change¹⁶.

The report stated that modern postgraduate curricula must require ‘demonstration of knowledge, skills and abilities through measurable and observable assessments’¹⁶. It found that there was an over-reliance on ‘time in service’ as evidence of progression and as a ‘proxy measure of competence’¹⁶. This was seen in the Annual Review of Competency Progression (ARCP) process, where trainee surgeons’ progress against the curriculum is measured on a yearly basis, and a summative decision is made whether to allow progress to the next year of training or implement some form of remediation.

The SoTR recognized that training must ‘continue to be bound by the timeframe of the training programmes’ and that extending training ‘will not necessarily lead to better trained doctors’¹⁶, but that training progress should be according to competencies and capabilities, and hence more bespoke to the individual. There was recognition that this might cause tension between ‘service continuity, delivery and training’¹⁶ but that ‘ultimately, it will give patients, doctors, trainers and employers more assurance that they have met the necessary requirements to work safely and competently with appropriate supervision’¹⁶.

The SoTR made two key recommendations for restructuring training based on the need for a broader and more flexible approach¹⁶;

- 1) That new graduates be fully registered with the GMC immediately after medical school (currently graduates are awarded provisional GMC

registration, which is upgraded to full registration upon successful completion of the F1 year)

- 2) That broad-based specialty training lasts 4-6 years after completing the foundation programme, and sub-specialisation occurs as a post CCST credential.

The reports' lead author, Professor Greenaway, summarized the findings of the review process; "in undertaking this review I discovered a wide recognition of the need for change" and a "clear consensus about what change should deliver: greater flexibility, better preparation for working in multi-professional teams and more generalists"¹⁶.

ii. Shape of Training Review Steering Group activity

In late 2013 Health ministers convened a UK-wide steering group (UK Shape of training group – the UKSTSG) to oversee implementation activity arising from the SoTR⁵⁰ and to undertake a detailed economic analysis of the proposed changes.

A pan-collegiate curriculum mapping review exercise was undertaken, to determine timescales for moving towards a more competency based training structure and to begin the transition to broad-based early training. The curriculum review exercise also sought to establish how future curricula should be quality assured and regulated (currently postgraduate curricula are updated by each respective Royal College and approved by the GMC who act as UK Regulator on all matters related to postgraduate training, since absorbing PMETB in 2010)⁵⁰. The considerable challenge the UKSTSG faced was how to apply these broad concepts in practice 'given the complexity of medical education and training and the parameters within which it was required to work'⁵⁰. The British Orthopaedic Association issued an early position statement setting out the potential implications of these in Trauma & Orthopaedics⁵¹.

At the same time as the initial work of the UKSTSG in 2013-15, the four UK Health Departments published their strategic plans for the future configuration and delivery of health services^{50 52 53}. Whilst somewhat changing the service delivery landscape

since the SoTR release⁵⁰, it was immediately clear that these plans share several common themes such as the need for greater flexibility in medical training and supporting broad-based practice, and thus could be implemented in tandem as collaborative ventures⁵⁰.

2018-2020: Improving Surgical Training Initiative

A new competency based training programme, the ‘Improving Surgical Training’ initiative⁵⁴ was announced as a joint pilot venture between The Royal College of Surgeons of England, Health Education England (HEE) and UKSTG in 2015. The pilot is being overseen by the Joint Committee on Surgical Training (JCST), which is an independent body that works on behalf of the four Surgical Royal Colleges of the UK on all matters related to surgical training^{50 55}.

i. The Improving Surgical Training vision

The IST pilot initiative aims to ‘create an improved surgical training system that produces competent professionals who are able to provide the highest quality of care to patients in the NHS’⁵⁶.

The key aims of the pilot are to⁵⁶;

- 1) Provide trainee surgeons with an appropriate balance between service and training
- 2) Professionalise trainers
- 3) Introduce a curriculum that is truly competence-based within a learning environment that embeds and enhances simulation
- 4) Ensure that the existing end-of-training product continues to meet current and future patient needs.

The potential benefits of this reform are wide-ranging. For patients, their surgeons will be trained to provide the highest standard of care, improving the continuity of care whilst also retaining access to subspecialist expertise where appropriate. For

trainee surgeons; there will be a greater quality and quantity of training provided by professional trainers, with a proposed reduction in duties of low educational value, and improved trainee/trainer relationship⁵⁶.

The key features of the IST pilot are;

- 1) Maximising training during daylight hours to ensure time spent on call is of educational benefit, including maximum on-call frequencies
- 2) Run through training with clear targets for progression
- 3) Support from non-medical professionals in delivering surgical care
- 4) Trainers undergo a programme of mandatory training and must demonstrate appropriate aptitude and qualifications to become educational supervisors
- 5) Trainers should be supported by an adequate amount of time in their job plans to manage, assess and support the trainee
- 6) Trainees should have a consistent relationship with a trained educational supervisor and a separate mentor
- 7) Each period of training should be long enough (minimum of 12 months) to allow trainers and trainees to develop a mutually helpful relationship
- 8) Trainee progression should be competence-based rather than time-based
- 9) Simulation should be embedded within the surgical curricula and there should be sufficient resource to ensure availability for all trainees
- 10) The early phase of training should be preceded by an educational induction where technical and non-technical skills are taught and developed in a simulated environment.

ii. Early concerns: is IST re-inventing the wheel?

A 2015 joint statement by the British Orthopaedic Trainees Association (BOTA) and the Association for Surgeons in Training (ASiT) expressed support for the IST consultation in attempting to improve surgical training, however several concerns were raised⁵⁷. These included strong opposition to a move towards generalism and any further shortening of specialist training⁵⁷. During the IST consultation process, the British Orthopaedic Association (BOA) and the Trauma & Orthopaedic

Specialist Advisory Committee (T&O SAC) chose not to join the IST pilot, as it was felt that the current T&O curriculum and CCT-holder end product already provided appropriately high quality care for NHS patients in the generality of Trauma & Orthopaedic' surgery. The IST pilot recruited its first cohort into general surgery in 2018, with a second recruitment round in 2019 expanding to include urology and vascular surgery.

iii. 2019 developments – unwelcome plans for expansion before review

Health Education England released a statement in July of 2019 saying that the IST pilot will expand to include 75% of core surgery posts by 2021, and 100% by 2022. It would also expand to include T&O⁵⁸. As of late 2019, there was not yet clarity on precisely when and to what extent T&O would be joining the pilot (this is based on a face-to-face conversation I had with the then-chair of the T&O SAC at a conference in September 2019).

The IST pilot is not due to report until late 2021, and concern has been raised from several stakeholders that premature expansion before the initial reporting phase is unwise⁵⁸. The 2007 events of MMC/MTAS are still in recent memory, and serve as a stark reminder of the potential consequences of reckless implementation of major training reform.

ASiT and BOTA released a position statement⁵⁸ in July 2019 in response to HEE's announcement of plans to widely implement IST before the success of the pilot had been analysed. They expressed concern over the delay in the independent review process, inadequate resourcing for the roll-out of IST, a dilution of the original objective and concerns for the surgical trainees themselves. The statement says 'IST is no longer a pilot and set to expand regardless of external review', and that 'premature expansion might negatively impact the engagement of stakeholders with detrimental consequences for the programme'⁵⁸.

Concern was expressed at the fact that 53% of the first intake of IST trainees have reported dissatisfaction in the balance between service provision and training, and 42% received protected or structured teaching from a consultant trainer once per

month or less⁵⁸, which is in direct opposition to the stated objectives of IST. They described evidence that dedicated time for teaching is set to be reduced by half and more anti-social hours work permitted, which suggests that the ‘original objectives and scope of the IST programme have been reduced in both scale and ambition’⁵⁸. They go on to say that ‘this rushed expansion represents a rebranding of early years surgical training without meaningful improvements in the delivery of training or the training environment’ and that ‘a once-in-a-generation opportunity to transform surgical training and prioritise training over service is being missed’⁵⁸.

ASiT and BOTA have requested the planned expansion of the IST programme to be halted with immediate effect, and that all stakeholders be given a role in future decision making and planning. They have requested that annual ‘waypoint reviews’ be implemented and shared with all stakeholders including the surgical colleges, and that these must enable the ‘objectives and scope of IST to be critiqued and amended as necessary’ so that IST ‘delivers tangible improvements in the quality of surgical training, with workforce morale, staff retention and patient care dependent on its success’⁵⁸.

Time will tell whether what appears to be an ‘IST-light’ version of the original mandate succeeds. At best, it could potentially represent an exciting new frontier in surgical education, applying creative solutions to the current problems with training in a struggling National Health Service. At worst, it could be a repeat of the MMC/MTAS fiasco.

Alongside the wider surgical education community, I look forward to the seeing the results of the IST pilot with interest, and a degree of trepidation.

1.1.2 The current challenges facing surgical training delivery in the NHS

The IST proposal was borne from a recognised need to improve and expedite surgical training following the SoTR, and to move away from time-based training, to

a competency-based model. It would be helpful here to discuss the recent drivers for this reform, as this sets the scene for the research presented in this thesis.

The NHS is under huge strain to meet patient demand

There is an unprecedented increase in the demand for healthcare in the UK across all sectors. The number of patients and care episodes is increasing year on year⁵⁹, and doctors are caring for more patients who are living longer with chronic illness and complex multi-morbidity⁶⁰.

A recent study estimated that 42% of the population of Scotland was living with at least one long term condition, and that 23% of the adult population had two or more long term health conditions⁶⁰. The fastest population growth will be those over 85 years of age, with significant associated increases in the associated medical and social care costs⁶¹. The number of people over 85 living in the UK is forecast to almost double in the next 25 years⁶².

The high demands on the NHS are clearly evidenced by declining performance figures for access to emergency departments, planned operations and ambulance response times⁶³, and the latest (2018) figures report a total NHS deficit of £558 million⁶⁴. The fact that solving the NHS crisis was central to all the main parties' political manifestos at the recent UK general election⁶⁵ in December 2019 further illustrates the gravity of the situation, as does the daily media coverage of the issues facing the NHS⁶⁶.

This rise in demand for services is set in a challenging climate of fiscal uncertainty and austerity, and of staff shortages driven by difficulties with recruitment and retention across all sectors in the NHS⁶⁷. In England there was a 28% increase in the number of Accident and Emergency attendances between 2012/13 and 2016/17, and in the same time period there was a 5.4% drop in the number of medical students attending UK universities⁶⁸.

The adverse financial climate is having a direct impact on training, and there is evidence that when services are under pressure, time and resources for education are the first to be sacrificed⁶⁹.

Traditional professional boundaries are changing

NHS England's Five Year Forward report⁵³ mandates action to safeguard and improve the NHS. Central to this is a patient-centred approach to the future delivery of healthcare, which should be available seven days a week, and accessible closer to home. The report has generated much debate about the way in which the NHS workforce plays a role in delivering this, and may involve blurring of traditional professional boundaries to provide a care delivery structure that is more responsive to patient need.

In surgery, this includes the recent introduction of non-medical healthcare professionals such as specialist nurse practitioners and physician associates/assistants, who take on some of the work traditionally performed by junior doctors. A 2016 report from the Royal College of Surgeons entitled 'A question of balance' examined the benefits of an extended surgical team for surgical training⁷⁰.

These benefits included the ability of non-medical professionals to take on some of the administrative and non-educational clinical service workload burden that now dominates much of the surgical trainees' time. This, when carefully managed, can render surgical trainees 'almost supernumerary' and free to pursue training opportunities in the operating theatre and outpatient clinics.

There is an emphasis on protecting and promoting patient safety

The Francis enquiry⁷¹ into the failings at the Mid Staffordshire NHS Foundation Trust between 2005 and 2009 focused public attention on patient safety, care quality and professional and organisational behaviours. The subsequent Berwick review into

patient safety⁷² catalysed a series of culture change reforms⁷³ including transparency, candour and freedom to speak about patient safety concerns⁷⁴.

These led to widespread efforts to improve staffing levels, support and supervision within healthcare teams, as well as stricter governance of doctors' training and practice through continued professional development and revalidation processes.

Specific to surgery, the culture of practice has changed to maximise patient safety. For example; the mandatory use of the WHO Surgical Safety checklist before every operation to reduce human error⁷⁵, defining and reporting 'never events'⁷⁶ (such as wrong site surgery), and rigorous audit activity of surgeon and departmental outcomes⁷⁷. In orthopaedic surgery, the National Joint Registry (NJR) collects data on all arthroplasty operations in the UK to monitor outcomes, identify failings and provide post-market surveillance of implants⁷⁸. There is an emphasis on transparency in assessing competencies and failures, and a culture of reflective, blame-free learning through critical incident reporting mechanisms within hospitals⁷⁹.

This welcome cultural change extends from everyday clinical practice through to surgical training. There are important patient safety implications in surgical training, beyond the obvious imperative that well-trained surgeons perform safer operations. The postgraduate training year in the UK begins on the first Wednesday of August, when newly qualified doctors start their first jobs and existing doctors move up a year and rotate to new placements. There is a well-documented, 6-8% rise in all-cause patient mortality in patients admitted on the August changeover day compared to a week earlier⁸⁰, the phenomena has been dubbed 'Black Wednesday' by the media⁸¹. There is longstanding concern that patients are exposed to mistakes by novice junior doctors, and that a modern training system should be able to mitigate risk to patients⁸².

It is no longer considered acceptable to allow trainee surgeons to 'practice' on patients⁸³, and it is probably reasonable to assume no patient wishes to be the recipient of a trainee surgeon's first attempt at a procedure. It has been stated by the Department of Health that simulation should be embedded in training⁸⁴, so that the early part of surgical learning curve can be moved away from patients, and a degree

of competence assured before the trainee is allowed to perform procedures on real patients.

There is a known quantity problem with surgical training

Today's surgical trainee is spending less time at work in the hospital, as a result of legislation that restricts their working hours to a maximum weekly average of 48 hours^{32 85}. It has been estimated that in previous generations the finished product of surgical training, a so-called 'day 1 consultant', would have accrued about 30,000 hours of operating time during training. That figure now stands at around 6,000 hours for the day 1 consultant in 2020⁴ – a five-fold reduction.

An independent review of the impact of the European Working Time Directive on postgraduate training was commissioned in 2010 by the Secretary of State and the resulting report, 'Time for Training', was written by Sir John Temple. The report found that the 15,000 hours available to trainees from a 48-hour working week on a typical 7-year specialty training programme were not being used effectively for training⁸⁶.

Reasons cited by Temple included rota gaps, poor rota design and shift patterns⁸⁶ which meant missed training opportunities in the elective, day-time setting. Temple found that specialties with a significant emergency care/out of hours workload – including orthopaedic surgery - were most affected by this. The report also found a large proportion of trainee practice was occurring with less senior supervision and hence missing valuable trainer-trainee interaction. This is compounded by a loss of continuity of care and disruptive shift-based work patterns, which affect trainee well-being and work-life balance⁸⁶. The increase in trainee numbers and increased scope of practice of Allied Health Professionals (to offset the reduced working hours and maintain service delivery), was reported to have 'diluted the quality and quantity of training posts'⁸⁶ and further threatens the training opportunities available.

More than 60% of consultants in a separate study reported concerns that the standards of surgical training would decrease within a 48-hour working week⁸⁷. Many consultant trainers' perception of how best to deliver training were 'aligned to traditional methods, involving long hours, personal sacrifices and learning, with limited formal educational support and supervision'⁸⁶, which is evidence of the pervasiveness of the apprenticeship-by-osmosis model described above.

The Temple report suggests that the reduced working hours can be used as a catalyst for training reform, and that high quality training can be achieved in a 48-hour working week⁸⁶. European comparators suggest that this is possible, as Denmark has had a 37-hour working week for surgeons⁸⁸ for many years and has not encountered the issues described by Temple.

Central to the Temple recommendations for maintaining training quality in the 48-hour reduced working week is to make 'every training moment count', where training must be 'planned, focused and individualised'⁸⁶. This includes using simulation to 'support and accelerate the learning curve'⁸⁶. It was 'consistently recognised' that the 'early acquisition of clinical skills should not be carried out on patients' which can pose a real threat to patient safety if 'trainees are required to deliver service before they have completed their basic training'⁸⁶.

Other workable recommendations from the Temple report include; 1) increased use of multi-disciplinary hospital at night teams (HaN), which maintain safe care for patients with less reliance on trainees, 2) implementation of a flexible, consultant delivered (rather than trainee delivered) service, which will increase the quantity and quality of training interactions, and 3) appropriate reward mechanisms to recognise excellence in training⁸⁶.

There is an emerging quality problem in surgical training

When a surgical trainee is at work, a very significant proportion of their time is taken up by activities that are administrative and offer no training value^{86 89}. The high

turnover of patients, increased pressure on discharges, shortage of support staff such as phlebotomists and pharmacy assistants, combined with inefficient bureaucratic processes (for example dated and slow IT systems and continued reliance on obsolete technology such as faxes and traditional pagers) means that an unacceptable amount of the trainee surgeons' time is consumed with administrative tasks⁷⁰. Whilst completion of these tasks is important to service delivery, they offer no training value and could easily be performed by someone with an administrative, rather than medically trained, background⁷⁰. This problem was starkly illustrated in a 2016 survey of 990 surgical trainees, which found that in an average 12-hour shift, trainees spent 218 minutes on administrative tasks, against only 34 minutes learning surgery in the operating theatre⁷⁰. In other words, the average surgical trainee, during the average shift, was spending more than 6 times the amount of time on administration than learning surgery. This, it is fair to say, is not 'making every moment count'.

The most valuable training environment for a trainee surgeon is without doubt the operating theatre and the outpatient clinic, access to which is typically only achieved once the ward and administrative workload is dealt with. There are many instances when this workload is insurmountable, such that the trainee is effectively stuck spending their time in ward-based activities that are not educationally valuable, and thus missing out on important training opportunities⁹⁰.

The overall exposure of the trainee to the operating theatre and outpatient clinics have decreased markedly since the introduction of the working hours restrictions^{29 91}⁹², and has been shown to reduce continuity of care and training opportunity²⁹.

A study by the Royal College of Surgeons showed that operative experience of basic procedures by trainees has dramatically declined in recent years, as measured by surgical logbook numbers⁵⁶. This logbook data shows there is very little operating performed in core training⁹³, and that UK surgical training is front loaded with excessive non-educational service provision at the expense of training⁹⁴. In other words, the average surgical trainee does not achieve meaningful exposure to operative training until their 5th postgraduate year (PGY). It is therefore not

surprising that satisfaction rates with training are the lowest amongst core surgical trainees⁵⁶, and the rates of attrition and burnout the highest in this group⁹⁵.

There is an increasingly significant problem with trainee retention and morale

In addition to the problems with both the quality and quantity of training described above, there are known issues with workload intensity for trainees, which are perhaps not surprising in a struggling health service. The annual national survey of doctors by the GMC reported that 64% of junior doctors* in surgery described their daily workload as 'heavy' or 'very heavy', as compared to an average of 40% of doctors in other specialties⁹⁶. 70% of trainers agreed⁶⁸. More than 80% of trainees** felt morale was declining⁶⁸. Surgery consistently underperforms relative to most other specialties for workload, satisfaction and burnout⁹⁵.

The link between staff and patient satisfaction is well established⁹⁷, yet a large majority of junior doctors report feeling undervalued, by the managers (88% of respondents), chief executive and organisation (77% of respondents) and the NHS as a whole (79% respondents). Worryingly, almost 60% did not feel valued by their consultants⁹⁸.

This poor morale among junior doctors is multifactorial. Pay freezes in the NHS (affecting all staff, including doctors), mean that since 2010 pay has fallen behind inflation⁹⁷. Trainees are leaving medical school with increasing debt⁹⁹, and the personal cost to individual trainees of achieving the mandatory requirements for completion of surgical training range between £20,000-£26,000 (dependent on specialty), compared to £2200 for medicine⁹⁹.

*The term 'junior doctor' is widely used to describe any pre-consultant level doctor. Much as I dislike the term, I have used it in this thesis in line with common parlance when referring to doctors of all specialties who are not yet consultants.

**When referring to surgical specialties only, I have sometimes used the term 'trainee', which I also think is demeaning and risks undermining the level of skill, experience and responsibility these people have. Surgeon-in-training is a better description, and I have tried to use this wherever possible.

Childcare that is sufficiently robust and comprehensive to accommodate the demands of an irregular and anti-social surgical rota is prohibitively expensive for many trainees, and is cited as a reason for the high levels of attrition from training – the so called ‘leaky pipeline’^{100 101}. This primarily affects female trainees^{102 103}, who carry a measurably greater average unpaid care load and domestic work burden than do their male trainee counterparts¹⁰⁴ – a further discussion of which is important but beyond the scope of this chapter.

There has also been a gradual erosion of ‘perks’ for junior doctors, including the loss of on-site accommodation for F1 doctors since 2002¹⁰⁵. This has been followed by widespread removal of doctors mess facilities and on-call rooms to make space for offices¹⁰⁶. Aside from the obvious practical aspects of losing these facilities (57% of junior doctors in one study reported they had experienced either an accident or near-miss when driving home from a night shift¹⁰⁷), these spaces represented a uniquely valuable place for debriefing, peer-support and socialising¹⁰⁸. This is well known to be an important coping mechanism in dealing with the psychological and emotional stressors of the job¹⁰⁸, and their loss is an important piece of the puzzle in discussing declining junior doctor morale¹⁰⁶.

Furthermore, the move to shift-based working systems has meant a loss of the traditional surgical ‘firm’. The support, mentorship and training opportunities the old ‘surgical family’¹⁰⁶ firm structure fostered has led to surgical trainees feeling disenfranchised and ‘just another number on a rota’¹⁰⁶. Surgical trainees will find themselves working with different senior (and junior) colleagues at every shift, and it becomes impossible to develop meaningful professional relationships under these transient conditions. The poor management of rota design has led to ‘petty tortures’¹⁰⁹ against trainees in trying to balance their lives – there have famously been instances of junior doctors denied leave to attend their own weddings¹¹⁰.

In addition to the problems with shift based working, the short nature of surgical rotations, sometimes as little as 4 months, mean that trainees are barely able to settle into their job role before being moved on. It is no surprise that permanent staff in these departments have little interest in getting to know trainee surgeons’ names, and the consultant trainers cannot possibly be expected to constructively mentor someone

they barely know. This rootlessness and sense of not belonging underpins much of the unhappiness seen in junior doctors¹¹¹.

These issues are also set within a wider feeling that the professional status of doctors (of all levels and specialties) has been gradually eroded^{112 113}. Cultural changes, including a ‘welcome growth in transparency and decline in patient deference’⁹⁷ has had the benefit of empowering patients and flattening the traditional medical hierarchy¹¹³. At the same time, the gradual creep of managerialism¹¹², and the increasing reliance on guidelines and protocols has been seen as ‘reducing the scope for legitimate clinical autonomy’ and undermining the professional status of clinicians⁹⁷.

The intensive regulation and assessment of doctors through workplace based assessments, continuing professional development and mandatory revalidation exercises has been widely cited as another important factor in the morale crisis, and are seen by many as burdensome, time consuming and heavy handed^{114 115}. The General Medical Council has been explicit in its recognition of this¹¹⁶, and has stated that it is committed to reducing the burden it places on doctors in its role as regulator⁶⁸.

The recent, high profile case¹¹⁷ of a junior doctor being convicted of gross manslaughter negligence following the death of a child in hospital brought both the patient safety consequences of junior doctors’ working conditions and the morale crisis to the forefront of public and media attention^{118 119}. In this particular case, the doctor was trying to do the work of several people due to chronic gaps on the rota, and was simultaneously trying to cover the clinical assessment unit, the emergency department and six wards spread over four floors. An IT problem meant blood test results were not available, and the supervising consultant had not realised they were supposed to be on call that day. Furthermore, this doctor had just returned from a year of maternity leave, and had received no induction or support in her return to work.

The ensuing viral success of the #IamHadiza hashtag¹²⁰ and outpouring of support from the profession by way of crowdfunding¹¹⁹ and petition campaigns¹²¹ was driven by the fact that most doctors recognised themselves as having previously worked in similarly dangerous situations, which had unfortunately led to an avoidably tragic outcome in this particular instance. The fact that a junior doctor was forced to work in impossible conditions and was then ‘singly scapegoated’ and tried in a criminal court for making an ‘inevitable mistake’¹²² was described as a ‘lightning rod for a profession with already rock-bottom morale’¹⁰⁶ and was met with a call for the resignation of the head of the GMC by the Hospital Consultants and Specialists Association¹²³.

As a result of these morale problems, an ever increasing number of junior doctors are choosing to take a break from formal training after completing the two year foundation programme¹²⁴, with over half of whom doing so because they feel burnt out after just two years as a doctor⁶⁸.

In 2010, 83% of foundation programme completers progressed on to speciality training. In 2018, just 38% did, and the number leaving medicine permanently had trebled¹²⁵. The application rates to surgical specialities has continued to fall year-on-year and trainees remain consistently less satisfied with their training experience than medical and general practice trainees¹²⁶.

The 2016 GMC National Training Survey results showed that trainee dissatisfaction had reached unprecedented levels^{97 127}, with the controversial contract reforms and widespread low trainee morale culminating in the first junior doctors strike in 40 years⁹⁸. This was the first ever all-out strike in the history of the NHS¹²⁸. A subsequent report commissioned by Health Education England¹²⁷ examined the reasons behind this, and found the key issues driving low morale to be burnout, lack of training opportunity, loss of team structure¹²⁸, lack of senior support and trainees not feeling valued by their employers¹²⁷.

The HEE report has been widely welcomed¹²⁷, and several projects are underway with key stakeholders to try to bring about meaningful improvements to junior doctors working lives. These exercises are still in the consultation phase and have

therefore yet to make their impact. Clearly, to ensure the future of the surgical profession, it is vital that surgery remains an attractive career to the brightest and best candidates.

The surgical workforce is changing

Another factor to consider in understanding the surgical training climate is the changing demographic of trainee surgeons. In 1991, just 3% of consultant surgeons were women, and in 2016 the proportion had risen to 11%¹²⁹. Trauma & Orthopaedics remains considerably behind the trend with just 5% of consultant surgeons being female in 2016, but the proportion of women at ST3 level is currently around 25%¹³⁰. This changing demographic is significant with regards to planning future training reform, as women surgical trainees are significantly more likely than their male counterparts to take time out of training to raise children, and to return to work on a less-than-full-time (LTFT) basis – 91% of LTFT trainees in 2017 were female⁹⁶.

Some trainee surgeons are also choosing to take time out of the clinical training programme to undertake a higher research degree, through formally integrated clinical academic training pathways¹³¹. These trainees, by virtue of their research endeavours, will take longer to progress through the training programme and considerations need to be made for their return to clinical practice following a prolonged period out-of-programme¹³².

1.1.3 Summary

I have described in this chapter how there is both a quantity and a quality problem with surgical training; a situation of reduced training time overall, with threats to the quality of this time in the form of loss of continuity of care, difficulty accessing the operating theatre and outpatient clinics on a background of excessive non-educational workloads and declining morale. The challenge is therefore to ensure that training quality is maintained within the constraints of the reduced hours and

service delivery pressures, and made flexible to accommodate the growing diversity of trainees. There is a clear need for new ways of training surgeons.

1.2 Thesis Aims and Objectives

1.2.1 Primary Aim

The primary aim of this thesis is to explore the effectiveness of a cadaveric simulation training (CST) intervention on the real-world surgical performance of junior orthopaedic surgeons-in-training.

1.2.2 Objectives

This thesis has three key objectives.

Objective 1: To make the case for using cadaveric simulation for postgraduate surgical training

This will answer the following questions;

- How do surgeons learn complex motor skills?
- What types of simulation can be used for training surgeons?
- What is so special about cadaveric simulation?
- What evidence is there for the use of cadaveric simulation for training postgraduate surgeons- in-training?
- What simulation is currently provided in training programmes for Trauma & Orthopaedic surgeons-in-training?

Chapter 2 contains a detailed discussion of educational theories around motor skill development, and a taxonomy of surgical simulation. Chapter 3 contains a systematic review of the current evidence for cadaveric simulation for postgraduate

surgeons-in-training. Chapter 4 reports the current status of simulation provision in the UK and Republic of Ireland (RoI) specialist training programmes in Trauma & Orthopaedics.

Objective 2: To explore how the technical skills of surgeons-in-training can be assessed

This will answer the following questions;

- What technical skills assessment tools are currently in use in Trauma & Orthopaedic surgical training?
- What is the evidence base for their use in practice?
- Can technical skills be assessed by measuring implant position on the post-operative radiographs?
- What measurements are relevant to both the technical skill of the surgeon and clinical outcome for the patient?

Chapter 5 is a systematic review of skill acquisition and operative competency assessment in Trauma & Orthopaedic training. Chapter 6 describes the development of an objective outcome measure of real-world technical operative success using post-operative radiographs.

Objective 3: To measure the impact of cadaveric simulation training on real-world surgical performance

This will answer the following questions;

- What are the early surgical skill acquisition trajectories of cadaveric simulation trained and standard trained junior Trauma & Orthopaedic surgeons-in-training?
- Do cadaveric simulation trained surgeons-in-training do better quality operations than standard trained-surgeons as measured by the post-operative radiograph?

- Do the patients whose operations were performed by cadaveric simulation trained surgeons have better clinical outcomes than patients whose operations were performed by standard trained surgeons?
- How are skills learnt in cadaveric simulation transferred into the operating theatre?
- What is it about cadaveric simulation that enhances learning?

Chapter 7 is a randomised controlled trial (RCT) of cadaveric simulation training versus standard on-the-job training. Chapter 8 is qualitative study of how cadaveric simulation influences learning in surgeons-in-training.

1.3 Potential implications of this work

I have described in detail the challenges facing delivery of surgical training in the UK, and the recent drive by training bodies to incorporate simulation into surgical training to help mitigate some of these challenges.

The work I present in this thesis will provide a clear overview of the educational theory supporting the use of simulation for surgical training. I will provide a systematic appraisal of the existing evidence for educational impact of cadaveric simulation in postgraduate surgical training. I will report on the current status of simulation provision in T&O training programmes in the UK and RoI.

I have systematically appraised the evidence for technical skill and operative competency assessment in T&O, and use the results of this to choose a suitable outcome measure that can objectively measure the real-world performance following a cadaveric simulation training intervention.

I have designed a randomised controlled trial to compare patient outcomes for cadaveric versus standard-trained surgeons-in-training. This will provide the first evidence of the impact of cadaveric simulation training on real world patients in

T&O. I have undertaken an in-depth qualitative study to provide evidence of how cadaveric simulation works.

My overall hope is that this body of work will contribute evidence towards policy decisions on centralised funding and universal simulation provision in T&O surgical training.

1.4 Conclusion

In this chapter I have described the surgical training landscape in the UK, the recent history of postgraduate training reform which has led to the current training structure, and I have discussed the issues facing trainees and trainers in the delivery of postgraduate training in T&O Surgery. This has set the scene for my research which I have stated in the thesis aims and objectives.

In the next chapter I will describe the emerging role of simulation for training surgeons and explore the educational theory supporting the use of simulation for surgical training.

Chapter 2: The emerging role of simulation for training surgeons

In this chapter I will explore the role of simulation to train orthopaedic surgeons. I have described in detail the current surgical training landscape in the UK and the challenges presently facing postgraduate training. In this chapter I will make the case for using simulation to address these challenges, with reference to the educational theory. I will describe the taxonomy of simulation in surgery, and cadaveric simulation, including the ethical and legal position of using cadavers to train surgeons.

Declarations

Parts of this chapter have been published;

James HK. Measuring the educational impact of simulation training in Trauma & Orthopaedics. *Journal of Trauma & Orthopaedics*. 2019;7(3):54-56.

James HK. Use of cadavers to train surgeons: what are the ethical issues? *J Med Ethics* Epub ahead of print: 2019;0:1-2 doi:10.1136/medethics-2019-105873.

Mrs Yessica Diez-Davies (YDD), professional graphic designer, improved the appearance of my original diagrams

Sections 2.5.1 and 2.5.2 (p. 74-76 only) are partly based on an assignment previously submitted for the award of the Masters in Medical Education

2.1 Introduction

A central tenet of healthcare delivery is *'primum non nocere'* – first do no harm. A 2019 meta-analysis of preventable patient harm in hospitals¹³³ found that worldwide, approximately 1 in 20 patients are exposed to preventable harm in healthcare. A focus on reducing the instance of avoidable harm in hospitals has been prioritised by the international patient safety policy agenda¹³³, but there are currently 'limited quality improvement practices specifically targeting incidents of preventable patient harm'. It is widely believed that development and implementation of strategies to specifically target preventable patient harm could lead to significant quality improvements and cost-savings in healthcare¹³³.

The reliance on the master-apprentice model of surgical training means that traditionally the early part of the surgical 'learning curve' took place on patients and might expose them to preventable harm from inexperienced surgical trainees. In striving to reduce the rate of preventable harm, attention should be focused on moving the early part of the surgical learning curve away from patients. Simulation-based education (SBE) is one way to do this.

Healthcare simulation, in a broad sense, is 'a technique to replace or amplify real-patient experiences with guided experiences, artificially contrived, that evokes or replicates substantial aspects of the real world in a fully interactive manner'¹³⁴. Crucially, simulation offers the opportunity for learning that is both immersive and experiential¹³⁵.

The potential gains from simulation-based healthcare are wide ranging. At an individual level, it can be used to improve skill level and demonstrate competence¹³⁶¹³⁷. At a team level, it can foster and improve interprofessional communication and teamworking¹³⁸¹³⁹. At a systems level it can improve overall performance of a service or department¹⁴⁰.

There are three broad domains for the application of simulation in healthcare¹⁴¹. These are; 1) simulation for the practice and assessment of technical procedures, 2)

simulated patients for the teaching and assessment of clinical skills and 3) simulation of complex scenarios for team-based training.

The work I have presented in this thesis relates to the first of these three; simulation for attaining technical competence. A discussion of the major motor skill acquisition theories is necessary to understand why simulation is a useful training tool in surgery¹⁴².

2.2 Motor skill acquisition theory

2.2.1 The neurobiology of the motor-skill learning curve

There are several established theories of motor skill acquisition in the literature, a detailed comparison of which is beyond the scope of this thesis. White¹⁴³ neatly synthesizes these various theories and describes two principles that are broadly common across the different approaches;

1) Changes in performance due to training and practice proceed in a non-linear manner.

This is the ‘learning curve’. Observed improvement will initially be rapid where there is large scope for improvement, and then decrease on repetition and continued practice as learning moves towards fine-tuning. Interestingly, the neurological processes that underpin skill acquisition appear to change as a trainee moves through the intermediate and advanced stages of learning^{144 145}. Functional neuroimaging has shown that in the novice, neural control when performing a new skilled action requires more ‘conscious, executive resources (e.g. pre-frontal cortex activity)’, whereas in the advanced learner recruitment activity is seen in more automatic centres of the brain (e.g. basal ganglia). Thus in the novice, rapid performance improvements are seen as neural programming takes place in response to exposure to the new ‘task arena’, whereas for the more advanced learner slower refinements of already-learnt processes take place over the course of repeated practice¹⁴³. This

accounts for the classic learning curve shape – initially steep during rapid skill acquisition, and then levelling off as refinement occurs.

- 2) As a trainee advances in perceptual-motor learning, the way that they process the task and action will change through qualitatively different phases.**

As proficiency improves through practice, the trainee moves from conscious, cognitive effort through to a state of automaticity, in which ‘the component subskills of a task can be performed in an effortless, offline manner’¹⁴³. Crucially, mental resources are freed up to focus on external goals and variations or disturbances in the practice environment. Hence once automaticity is achieved, the learner can adapt their performance efforts towards ‘conscious goals and evolving task restraints’¹⁴³.

These stages were first described by Fitts and Posner in 1967¹⁴⁶ (figure 1).

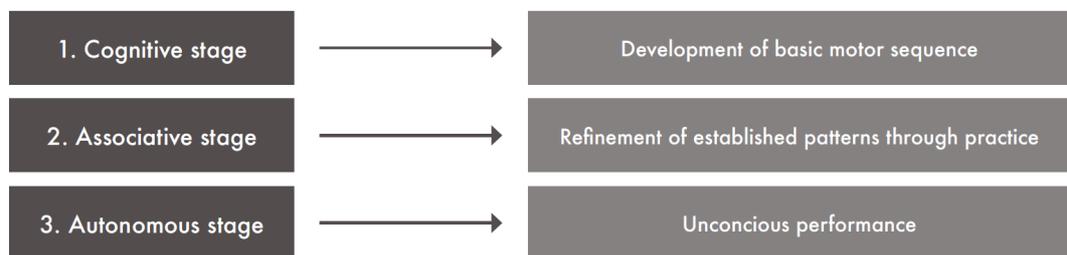


Figure 1. Fitts and Posner’s three stages of motor skill acquisition

Building on this simple model, Dreyfus and Dreyfus proposed an influential framework by which learners acquire skills and progress from ‘novice’ through to ‘expert’ by a combination of formal instruction and practice. This framework was originally reported in their 1980 research thesis describing the skill development of military pilots at the Operations Research Centre of the US Airforce¹⁴⁷, but has since become widely cited across the medical education literature, and accepted as a central tenet in understanding the learning curve of surgeons¹⁴⁸.

Stage	Knowledge application	Relevance recognition	Assessment of context	Decision making
Novice	Without reference to context	None	Analytically	Rational
Advanced Beginner	In context			
Competent				
Proficient		Present	Holistically	
Expert				

Figure 2. Dreyfus and Dreyfus model, adapted from <http://www.sld.demon.co.uk/dreyfus>

2.2.2 Surgical procedures as perceptual-motor skills

All surgical operations are comprised of a series of variably complex motor tasks. These motor tasks can be deconstructed into functionally individual psychomotor skills¹⁴⁹. Psychomotor skills in the context of surgery can be defined as ‘a stable and reliable link between perception of body and environment, and execution of goal-directed motor actions, which is both consistent across repeated performances of the action, and can be flexibly adapted to changes in task constraints’¹⁴³.

Surgical skills are considered a primarily **perceptual-motor** skill similar to instrumental music and athletic performance, in contrast to more broadly **intellectual** skills such as chess playing or mathematical deduction¹⁴³. Successful execution of a perceptual-motor skill, and therefore surgical performance, involves fluent integration of visual and haptic perception, fine physical movement, temporal and spatial awareness and tool manipulation¹⁴³.

In teaching complex motor tasks such as surgical operations, the practice or simulated exercise can be arranged as either ‘part practice’ – the task is broken down into its component skills which are taught (and assessed) in isolation, or ‘whole practice’ – the component skills are taught in their entirety, in their naturally

occurring serial order¹⁴⁹. Briggs and Naylor¹⁴⁹, early pioneers of motor skill acquisition theory, theorised that a task's complexity and its organisation determine the optimum type of practice¹⁵⁰. **Task complexity** is defined as the number of movement segments, and **task organisation** is the temporal relationship between the composite movement segments¹⁴⁹. Surgical operations can therefore be considered as a series of high complexity, low organisation tasks – in other words the movements required are functionally complex but they occur in a fairly predictable sequence. According to Briggs and Naylor, optimum learning of such tasks occurs during **part-practice** conditions^{149 150}. The opposing view is that improvements in motor task performance can be considered as the performance of individual skills together in sequence such that the transitions between the individual skill components disappear¹⁴⁹, as such **whole-practice** may be better to achieve task fluency. Several studies suggest whole-practice¹⁵¹, or combined part-whole practice¹⁵² to be superior for learning complex tasks such as surgical procedures.

A further principle in the motor skill acquisition literature relates to **random practice** or **block practice**. In random practice, multiple procedural steps of a given task are practiced in a single session in a random order¹⁵³. By contrast, blocked practice requires skills to be practiced in consecutive, logically ordered blocks, with progression to the next 'block' occurring after a pre-defined amount of practice¹⁵³. Random practice increases the level of difficulty during initial skill learning, which may be reflected in lower initial post-test scores, but may actually serve to increase performance on retention and transfer tests¹⁵³.

By way of real life example, an orthopaedic operation that is commonly performed by junior surgical trainees is one that is done to fix a fractured (broken) ankle using metal screws and plates. This operation would typically involve the following **tasks**; prepping/draping the patient, soft tissue incision and dissection, fixing the fracture, haemostasis (stopping bleeding) and finally wound closure. Each one of these component tasks involves several individual perceptual-motor **skills**. In this example, the 'fixing the fracture' task could be broken down into the following six skills; reduction of the fracture, application of the metal plate, drilling, depth measurement, bone tapping and insertion of the screws¹⁴⁹.

According to the Fitts and Posner model, a junior trainee at the cognitive stage of motor skill acquisition will understand the skills required to complete the task at hand, and will perform erratically with discrete, distinct steps, requiring much concentration and thought¹⁴². In the example of the ankle fracture operation, they will expend a great deal of cognitive effort on holding and manipulating the instruments and undertaking each step of the operation; they may, for example, take numerous attempts to reduce the fracture, and to select and place the plate correctly. With appropriate feedback and practice (across an as-yet undefined number of training exposures) the trainee moves to the associative stage where they do not have to concentrate so hard on each aspect of the procedure and performance becomes more fluid. They reduce the fracture quite easily and apply the plate and screws without having to think too much about how to handle the instruments, for example. Further feedback and practice exposure eventually leads to achievement of the autonomous stage, where the trainee surgeon no longer needs to concentrate directly on motor behaviour, and can instead direct their attention to other aspects of the procedure – for example a challenging or unusual fracture pattern, intra-operative complications and interruptions and questions from colleagues.

In terms of designing the ideal surgical training intervention therefore, consideration should be given to; 1) the nature of the task being taught, 2) the component perceptual-motor skills and 3) optimal practice sequencing.

2.2.3 Deliberate practice for surgery

In the Dreyfus model (figure 2) learners progress through practice and instruction via five stages; novice, advanced beginner, competent, proficient and expert. The concept of ‘deliberate practice’, described by Ericsson¹⁵⁴, has helped elucidate the acquisition of expertise in surgery and enhance progression through the five Dreyfus stages.

In a landmark paper, Reznick et al.,⁷ emphasized that the total number of hours spent in deliberate practice of surgery, rather just being present in the operating theatre, is important in determining the eventual level of expertise attained by the learner.

Sustained deliberate practice has been shown to be essential for the development of expertise across a variety of domains, including chess, instrumental music performance and sports^{142 154 155}. In surgery, Ericsson defines ‘expertise’ as experienced surgeons with consistently better outcomes than non-experts, and states that volume of experience alone does not account for observed differences in skill level amongst surgeons⁷.

Deliberate practice involves⁷;

- 1) Repetitive performance of intended cognitive and psychomotor skills in a focused domain
- 2) Rigorous skills assessment
- 3) Specific informative feedback

Deliberate practice is therefore felt to be a critical process for the development of expertise or mastery⁷. Extensive research on human performance has shown that superior task performance does not automatically develop from repeated experience or general teaching, and that superior performance requires deliberate practice behaviour to actively set new goals and achieve higher performance standards. Studies into expertise and deliberate practice have shown that most professionals reach a stable, average level of performance¹⁴² and that around 10,000 hours of deliberate practice is required to achieve expert status in complex tasks such as surgery^{154 155}.

There is also an emerging body of evidence showing that mastery is not a permanent, entirely stable state as once believed, and that there is measurable skill attrition towards the end of surgical careers, thought to be in part due to a subtle age-related decline in cognitive and psychomotor functions¹⁵⁶⁻¹⁵⁸ – a reverse learning curve as it were. This is a fascinating area of research but beyond the scope of this chapter.

The implication of deliberate practice theory for surgical training is that whilst the required hours to achieve expertise are theoretically available within a typical postgraduate surgical training programme (assuming a 48 hour working week across 7 or 8 years), these hours are under threat from other factors such as an increasingly

large administrative and service provision burden (please see chapter 1). Hence current training conditions do not necessarily provide opportunity for deliberate practice with the appropriate feedback. A key solution to this is the use of simulation – to deliver optimal deliberate practice and motor skill learning conditions, and, crucially, to move the early part of the learning curve away from the patient.

2.2.4 Simulation enhanced learning trajectory

Reznick⁷ argues that basic skills should be taught to the autonomous stage using simulation, then learners can refine and build up complex skills in the operating theatre, and hence, in the words of MacCaskie¹⁵⁹, there is ‘no learning curve on patients’. As described in chapter 1, the culture and politics around healthcare delivery have changed significantly and safety concerns along with workplace time pressures mean that live patients are no longer an acceptable model on which to practice the acquisition of new skills¹⁴².

The need to streamline training and mitigate potential patient safety issues is central to the argument to incorporate simulation into training for surgeons.

Curran⁸⁴ expands on the original Dreyfus model, and proposes a ‘simulation enhanced’ learning trajectory, whereby the learner progresses more rapidly through the five stages by experiencing simulation, beginning with task simulation at its most basic level for novices, through to supervised clinical practice and mission rehearsal to augment the transition from proficiency to expertise. Curran’s model suggests that different types of simulation are most beneficial at different stages of the learning trajectory, and that the learner should initially acquire basic skills ‘*ex vivo*’ on simulators before achieving competence, at which point they reach a ‘threshold to practice’ and can begin supervised clinical procedures on live patients to refine skills towards a state of proficiency and subsequent expertise.

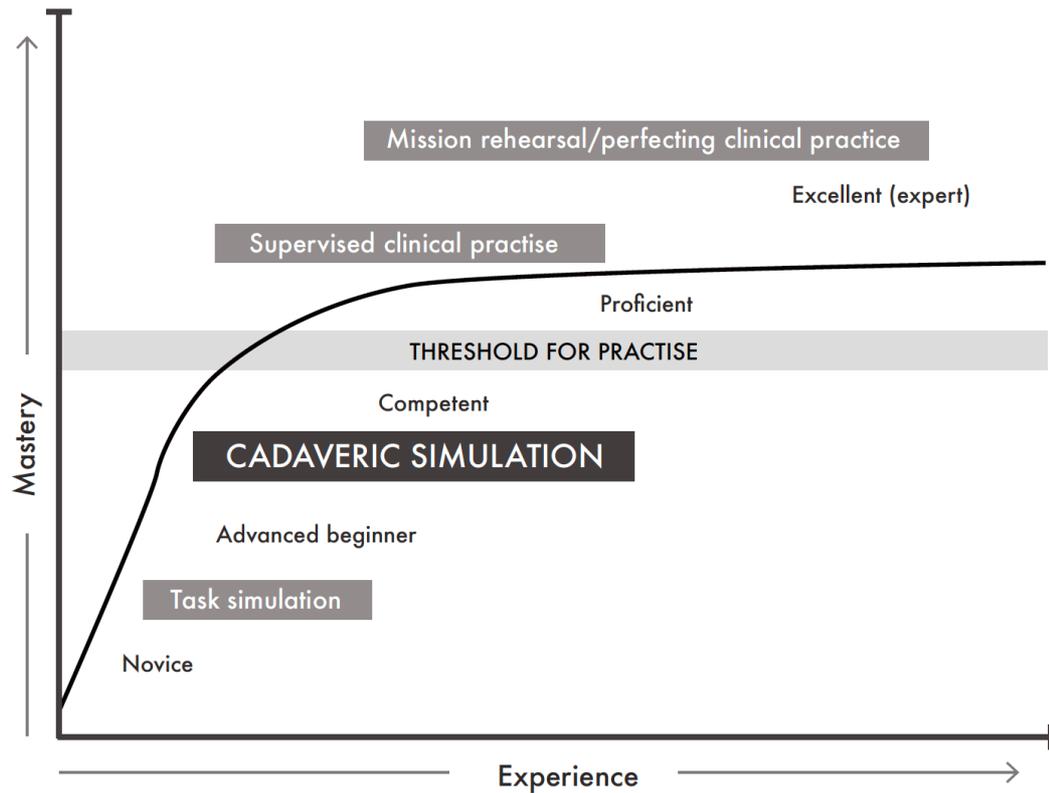


Figure 3. Curran's simulation enhanced learning trajectory⁸⁴

In this thesis I am focusing on the learning curves of very junior surgeons-in-training performing relatively simple procedures. The same principles apply, however, to experienced surgeons learning new techniques. Areas such as robotics and endoluminal surgery have their own learning curves and given the relative infrequency with which these procedures are performed, these could be quite long¹⁶⁰¹⁶¹. It has been said that simulation based training should be part of the set-up package of new surgical technologies to safeguard patients¹³⁵.

The widespread and rapid dissemination of laparoscopic surgery in the 1990s without any pre-adoption training package in place (and not enough 'master' surgeons to teach under the master-apprentice model) meant that there was avoidable patient morbidity and mortality¹³⁵, which has been described as 'the biggest unaudited free-for all in surgery'¹³⁵.

2.2.5 The increasing move towards simulation for surgical training

The movement towards using simulation to train doctors, and specifically surgeons, is rapidly gaining momentum in the UK. The Government, in several reports^{84 162 163}, have stated that developing and integrating simulation into surgical training is a strategic priority, in response to the challenges with training delivery described in chapter 1. The Chief Medical Officer has previously stated the potential safety benefits for patients offered by simulation and recommends that it be ‘fully funded and integrated within training for clinicians at all levels’¹⁶².

In the General Medical Council’s 2011 document entitled ‘The Trainee Doctor’, item 8.7 requires that ‘trainees must be enabled to improve their clinical and practical skills, through technology enhanced learning opportunities such as clinical skills laboratories, wet labs and simulated environments’¹⁶⁴. The Department of Health framework for Technology Enhanced Learning⁸⁴ reinforces this by stating that ‘healthcare professionals should learn skills in a simulation environment before undertaking them in supervised clinical practice’. The report goes on to make specific recommendations;

- "The use of simulation...should be achievable and clearly mapped to specific learning outcomes in identified areas of the curriculum" (recommendation 5c).
- "Healthcare, social care and education partners should aspire to educational excellence by encouraging innovation, evaluation and the dissemination and adoption of evidence-based, good practice" (recommendation 6).

The Chief Medical Officer has identified that there is an urgent need to build on the current evidence base for using simulation in training, to evaluate how this approach to learning can continuously improve patient care⁸⁴. More recently the Improving Surgical Training pilot has placed simulation delivery at the centre of its philosophy on training reform⁵⁶.

2.3 Taxonomy of simulation in surgery

It is clear that simulation training for surgeons offers great potential in expediting learning in a safe manner away from patients, in conditions that are underpinned by educational theory, and the move towards this type of training has considerable support from the Government and regulatory training bodies. What is not well understood is how trainees learn from simulation and how best to deliver this type of training. Within exploring how best to deliver simulation training, attention needs to be paid to both the type of simulation intervention, and the timing of its delivery within the postgraduate surgical training programme.

A brief summary of the types of simulation, and the features of simulation training that are already known to be beneficial to learning is presented.

A simple definition of simulation is ‘a reproduction or approximation of a real event, process and set of conditions or problems’⁸⁴, whereby the learner is expected to behave as they would in the real situation. Simulation is often described in terms of ‘fidelity’, which describes how realistic it is, in other words how accurately the simulation resembles the actual situation that is being reproduced⁸⁴. Simulation can be high or low fidelity (i.e. high or low realism) and can also be described more specifically in terms of environmental fidelity, equipment fidelity and psychological fidelity. Environmental fidelity is a measure of how accurately the simulation recreates the environment in which the event/procedure would occur in real life. Equipment fidelity similarly describes how realistic the equipment used for the simulation is as compared to real life. Psychological fidelity describes how the simulation reproduces the emotional and behavioural aspects of the real situation. Simulated scenarios that have high environmental, equipment and psychological fidelity are also known as ‘full immersion’ simulation⁸⁴.

An important factor when considering the use of simulation for training surgeons is what degree of environmental, equipment and psychological fidelity is required to achieve the educational objective in question.

Sarker and Patel propose a simple taxonomy to classify simulation types in surgery¹⁶⁵

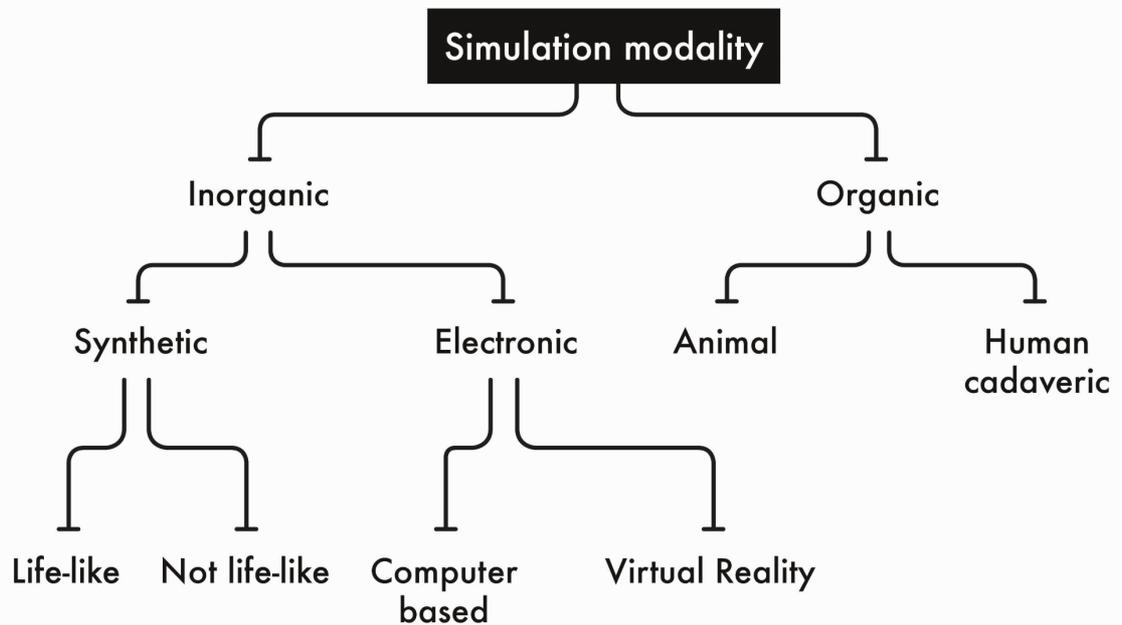


Figure 4. Taxonomy of simulation

Inorganic simulators represent a broad category, ranging from simple, low fidelity, low-cost bench models such as a simple suture jig¹³⁷, to moderate fidelity laparoscopic box trainers¹⁶⁶, to high fidelity, sophisticated virtual reality type simulators with the capacity to provide haptic feedback to the learner¹⁶⁷. Inorganic simulators as a group have the advantage of being ‘clean’, not requiring specialist maintenance or facilities for their use, and being cost-effective, representing a largely one-off financial investment for the host institution^{165 168}. Electronic simulators such as virtual reality increasingly offer the additional ability to capture data on learner performance (e.g. hand motion analysis, time to completion of task, task error etc.) which have been shown to be an important feature of their educational impact¹⁶⁹⁻¹⁷³.

Inorganic simulation is particularly effective for teaching laparoscopic skills and as such is used predominantly within specialties such as general and colorectal surgery,

where laparoscopic procedures represent a significant proportion of the typical surgeons' clinical workload^{167 174 175}.

The main disadvantage of inorganic simulation is that it does not allow for the realistic representation of anatomy and soft tissue handling that is fundamental to the practice of open surgery^{7 176}, other than developing skill in the most basic of open surgical skills such as extra-corporeal knot tying¹⁷⁴. Within orthopaedic surgery, which uses predominantly open techniques (with the notable exception of arthroscopy), there may be a role for using inorganic plastic bone models to teach novice trainees the basic principles of instrument handling and drilling technique¹⁷⁷.

Organic simulators, which use either animal or human tissue, can be used in open surgical simulation training. The use of live anaesthetised animals raises obvious ethical concerns and is currently not permitted under UK law, but is allowed in some circumstances in the US and Europe¹⁶⁵. Live animals can replicate bleeding but their anatomy is fundamentally different to that of humans, thus their utility in surgical skills training is probably limited to specific applications such as dealing with haemorrhage in the context of trauma^{165 178}, for example in military field training exercises¹⁷⁹.

Human cadavers have been permitted for use in surgical training since the introduction of the Human Tissue Act in 2004¹⁸⁰. Cadaveric simulation training (CST) has rapidly gained popularity in the ensuing years as an emerging simulation technology¹⁷⁶, with many UK teaching hospitals and medical schools now holding the required Human Tissue Authority license¹⁸¹.

Human cadavers are perceived to be educationally useful as they represent true anatomy as encountered in real life '*in vivo*', including physiological variants between individuals^{7 176 182}, and complex three-dimensional neurovascular relationships which are difficult to appreciate from textbooks and almost impossible to replicate in synthetic models¹⁸³. Importantly, cadavers offer the opportunity of being able to practice an operation in its entirety, from initial incision and approach through to wound closure, with high environmental, equipment, and with the right

design, psychological fidelity - therefore enabling advanced procedural skills to be developed⁷.

In spite of the numerous advantages of CST, there are some challenges with its use that have drawn criticism. Human cadaveric material is expensive¹⁷⁶, not-reusable, difficult to store and preserve, and requires specialist facilities and personnel for its use in training. There is also a theoretical risk of disease transmission. The fidelity of the simulation has been questioned as the tissues do not bleed in response to iatrogenic trauma as they do in life⁷, although this issue has been largely overcome by innovative efforts to perfuse^{184 185} (and even ventilate¹⁸⁶) cadavers. Effective cadaveric tissue perfusion has been demonstrated with angiography techniques and notable bleeding vessels at the skin margins, demonstrating that vascular reconstitution of cadaveric tissue can be achieved. This can maximise simulation fidelity for procedures where bleeding is an important part of the training exercise¹⁸².

The high cost associated with cadaveric simulation training can be mitigated by planning training sessions carefully around the short useful life of a cadaver, by performing the maximum number of possible procedures on each specimen. Blaschko and colleagues¹⁸⁷ have described a method for the successful co-ordinated use of shared cadavers between specialties to maximise the number of surgical procedure training sessions performed on each cadaveric specimen. Their results showed that ‘cadaver re-use increases the availability of cadaveric material without compromising tissue quality for dissection’, and that this technique ‘effectively halves’ the cost of using cadavers to train surgeons¹⁸⁷.

2.4 Educational theory supporting the use of simulation for training surgeons

Based on a qualitative synthesis of evidence in their BEME (Best Evidence Medical Education) systematic review on the features and uses of high-fidelity medical simulations that lead to effective learning, Issenberg and colleagues¹⁸⁸ describe ten key features of a medical simulation intervention. These are the features which best

facilitate learning, and which represent a set of goals to maximise the educational impact of simulation training.

These are summarised below;

FEATURE	
1	Providing feedback
2	Repetitive practice
3	Curriculum integration
4	Range of difficulty level
5	Multiple learning strategies
6	Capturing of clinical variation
7	Controlled environment
8	Individualized learning
9	Defined outcomes
10	Simulator validity

Figure 5. Issenberg's 10 features of effective simulation

It follows therefore, that the 'ideal' simulation would incorporate all 10 of these features in its design.

Meller¹⁸⁹ proposed a formal typology of medical simulators based on their component parts and participants. This typology allows the trainer, in designing an intervention, to make a judgement about the fidelity of the simulation needed to achieve its educational goal – the more component parts from the typology that are included in a simulator, the higher the fidelity.

COMPONENT	DESCRIPTION
P1	Patient or disease process of interest
P2	Procedure/diagnostic test/equipment being used
P3	Healthcare professional (usually the learner)
P4	Expert practitioner (usually the trainer)
P	Passive element of the simulator
A	Active element of the simulator
1	Interactive element of the simulator

Figure 6. Meller’s typology of medical education simulators

In addition to this summary of the philosophy of simulator design, brief attention to the current thinking around how to test learning on simulators is required.

Miller describes four distinct levels at which a medical learner should be assessed in practice, with reference to the stage of learning¹⁹⁰. Miller’s ‘triangle’ was first described in 1990 and has now become an accepted foundation principle in medical education research.

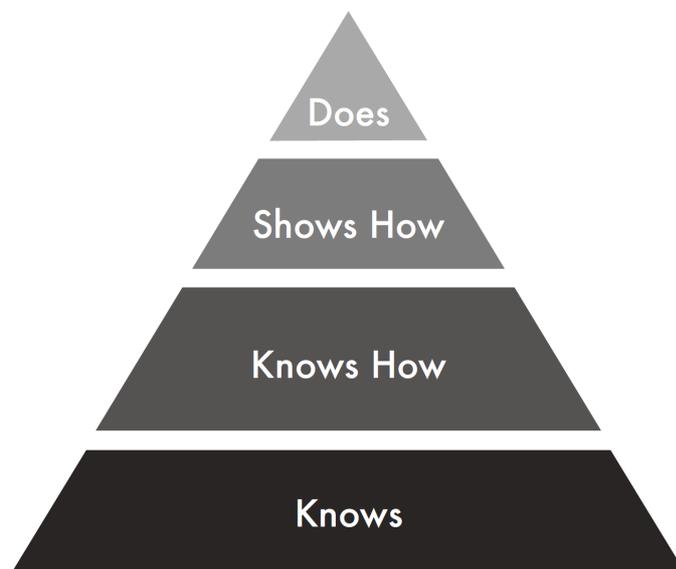


Figure 7. Modified Miller’s triangle

From the perspective of assessing learning, these four stages correspond to knowledge testing (knows), demonstration of competence (knows how), demonstration of performance (shows how) and action in clinical practice (does). Simulation facilitates learning at the top two levels – ‘shows how’ and ‘does’¹⁹⁰.

Kirkpatrick described a model for evaluating the impact of a training intervention on learners¹⁹¹, which is widely known as the ‘Kirkpatrick’s hierarchy’ in the medical education research literature.

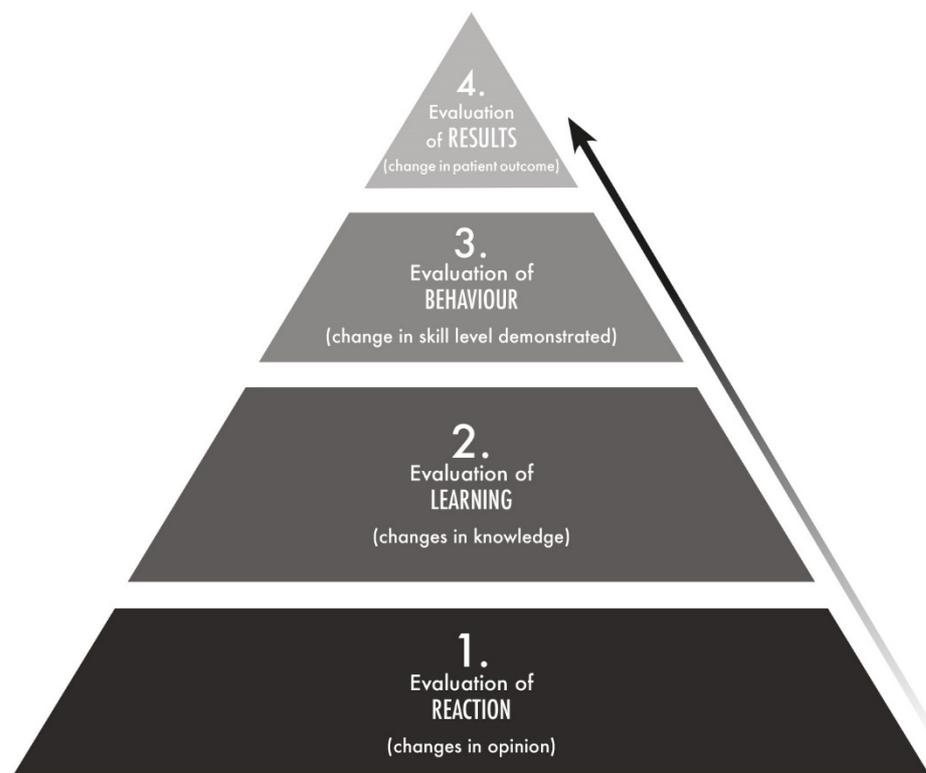


Figure 8. Adapted Kirkpatrick’s hierarchy

The original stages are shown in bold and their application to medical education theory given beneath, so what is widely cited in the literature as ‘Kirkpatrick’s hierarchy’ should actually be more accurately described as an adapted or modified version of the original.

From an educational theory perspective therefore, the best possible simulation training intervention would contain all 10 features of effectiveness as described by

Issenberg¹⁸⁸, the necessary constituent components to achieve the appropriate fidelity profile as described by Mellor¹⁸⁹, and the intervention would advance the learning to the ‘does’ stage described by Miller¹⁹⁰, with a resulting, demonstrable benefit to patients (level 4 of Kirkpatrick’s hierarchy)¹⁹¹.

From a practical and research perspective, consideration also needs to be given to the cost of the simulation training, the ethics of delivering the training, and as mentioned above, appropriate timing within postgraduate training to maximise benefit.

2.4.1 The evidence for the transfer of surgical skills into clinical practice following simulation based training

A theoretical appreciation of why simulation might be an effective tool for training surgeons is arguably worthless without empirical evidence to show skill retention and transfer into real-life clinical practice, and certainly does not in itself support major investment decisions and curriculum reform such as that necessary to truly integrate simulation into surgical training.

Two systematic reviews have attempted to synthesise the available evidence for skill transfer to the operating theatre following simulation training. The first, by Buckley et al¹⁹² reports the results of 16 studies, all randomised controlled trials, all concerning laparoscopic surgical procedures where the simulation intervention was delivered using video/computerised/virtual reality simulators. Of these, 11 of 16 studies measured operative time as the primary outcome measure. Eight of 16 studies used OSATS (objective structured assessment of technical skills in surgery) to measure performance differences between the simulator trained and control groups. Eight of the studies also used some kind of performance indicator as an outcome measure, five of eight used an error scoring system, and three of eight used a performance scoring system based on a formula incorporating time and accuracy. Seven of the 11 studies measuring procedure time (PT) as the primary outcome measure showed ‘significantly’ shorter operating time in the study participants who underwent procedural simulator training. All the studies that used OSATS as their primary outcome measure showed improvement in the scores after simulation

training relative to the control groups, and this was statistically significant in 88% of cases.

Of the studies using a performance indicator as the primary outcome measure, five of five found the error rate was lower amongst the surgeons who had undergone simulation training, and three of three using a performance scoring system showed improvement in the simulation trained groups. Importantly, in nine of 16 studies, skill transferability was measured in a porcine model, rather than the real operating theatre with human patients. The review team concluded that the current literature ‘consistently demonstrates the positive impact of simulation training on operative time and predefined performance scores’, however the authors recognise that these ‘reproducible measures alone are not sufficient to demonstrate transferability of skills from the laboratory to the operating room’¹⁹².

Within the reported studies, there were significant problems with poor methodology, reporting and small sample sizes. The outcome measures used to measure skill transfer were variable and not necessarily validated¹⁹². Operation time, for example, used in 11 of the 16 studies, may be relevant from an economic perspective but ‘does not reflect operative proficiency when measured in isolation’¹⁹², or indeed describe the quality of the end product of the operation. I will consider the utility of technical skills assessment tools in measuring proficiency separately in detail in chapter 5 of this thesis.

The second systematic review to assess skills transfer after surgical simulation based training, by Dawe et al¹⁹³, included 34 studies, of which 27 were randomised controlled trials, and seven were non-randomised comparative studies. Fourteen of these involved laparoscopic procedures, 13 involved endoscopic procedures and four were classified as “other”. Of these, one was an open surgical procedure (closure of abdominal wall fascia), using a synthetic, low fidelity simulation model¹⁹⁴.

The outcome measures used to assess skill transfer were operating time (18 of 34 studies), success rate (13 of 34 studies), performance errors (seven of 34 studies) and patient discomfort (four of 34 studies). Of the 18 studies that measured performance time, eight studies showed a significant reduction in operating time in the simulation

trained group. Of the 13 studies that measured success rate (defined variously as the percentage of surgeon participants able to complete the patient based assessment, the percentage able to complete the operation independently without assistance and the number of participants awarded a 'pass grade'), seven of 13 studies showed improvement in success rates following simulation training. Of the seven studies that measured performance error to assess skill transfer (defined as 'movements or events outside the normal procedure'), six studies showed that the simulation trained participants 'made significantly fewer intra-operative errors, and one study also found that simulation training was associated with lower intra and post-operative complications'¹⁹³. Of four studies which measured patient discomfort during the assessment procedure, one study reported an improvement in the simulation trained group (although this was not statistically significant), and three studies found no difference.

Dawe et al concluded that the review 'supports the hypothesis that simulation based training has advantages over no training', and that 'of 28 studies that made this comparison, 23 showed performance improvement in the simulation trained group and five found no difference', and that the 'evidence for the value of simulation training compared with patient-based training is weaker'¹⁹³. It appears (in the absence of a statement to the contrary) that the measures of transferability in the Dawe review studies were made during real operations on human patients, which arguably increases the strength of the findings.

Of note, between these two reviews, only one study involved an open surgical procedure, and none used cadaveric simulation as the training intervention.

2.5 The case for cadaveric simulation in surgery

2.5.1 The history of training surgeons using cadavers

The use of human cadavers to train surgeons is not a new phenomenon. Cadaveric dissection has played a major, invaluable role in the development of our modern

understanding of anatomy and physiology, and has long been recognized as a valuable training tool.

“He must mangle the living if he has not first operated on the dead”

Sir Astley Cooper (1768-1841), London Surgeon-Anatomist¹⁹⁵

Dissection of the dead for the purposes of understanding anatomy began in Alexandrian Egypt around 300BC¹⁹⁵, when bodies were routinely eviscerated as a part of the mummification process, which afforded the opportunity to study internal anatomy. In Europe, cadaveric dissection became popular in the eighteenth and nineteenth centuries with the advent of private anatomy schools, but until the passage of the Anatomy Act in 1832, the only legally available cadavers in the UK were the bodies of those executed by capital punishment, as they were the property of the state¹⁹⁵. Public acceptance of cadaveric dissection did not keep pace with the rapid expansion in the number of anatomy schools – the religious 18th century popular belief at the time was that dismemberment of the body after death jeopardized ones’ chances of resurrection¹⁹⁵, and dissection after death remained highly stigmatized and synonymous with capital punishment¹⁹⁶.

Early anatomists engaged in somewhat extreme measures to ensure a constant supply of fresh cadaveric material for dissection purposes, with body-snatching proving to be a lucrative business in early Victorian London¹⁹⁷. The passage of the Anatomy Act of 1832 was a pivotal period in the history of cadaveric dissection, as it excluded the use of executed criminals for dissection purposes and instead allowed for body donations. In actual terms, this meant that bodies were overwhelmingly sourced from the poorest and most vulnerable members of society, as they were sold by relatives to anatomists after death and also sourced from unclaimed deaths in institutions such as workhouses and psychiatric hospitals¹⁹⁶. Therefore, whilst solving the problem of body-snatching and grave-robbing, and ensuring a ready supply of inexpensive and fresh cadaveric material to medical schools, the science of anatomy remained ‘morally corrupt in its exploitation of the most vulnerable members of society for the purposes of scientific endeavour’¹⁹⁶.

The 1984 Anatomy Act aimed to simplify the process of body donation, and clarify the process of bequests made in life. Whilst undoubtedly welcome and necessary from a governance perspective, the 1984 act also restricted the use of cadavers to the exploration and definition of topographical anatomy, and forbade their use for surgical training purposes¹⁹⁸.

2.5.2 Legal position regarding the use of cadavers for surgical training

Recent legislative change has served to clarify the legal position with regards to the use of cadavers specifically for surgical training purposes. Alongside the Anatomy Act 1984, the Human Tissue Act 1961 governed the use of cadaveric tissues for diagnostic, research and post-mortem examination purposes, but did not give explicit guidance on their use in surgical training. The Human Tissue Act 2004 integrated and replaced these two separate pieces of legislation¹⁹⁸, and established the Human Tissue Authority (HTA), whose remit is to regulate the use of human tissue (including cadavers) in the UK. The HTA 2004 has ‘5 licensable sectors’; teaching, anatomy, research, public display and the making of a post mortem examination¹⁸⁰.

Within the HTA, cadaveric surgical simulation is considered as ‘education or training related to human health’ (Schedule 1, Part 2, Section 9) and is lawfully permitted if done with appropriate consent (Part 1, Section 1.1)¹⁹⁹. Cadaveric simulation training is considered a ‘scheduled purpose’ under the HTA rules, and therefore may take place at any UK hospital that has a HTA post mortem license and Designated Individual to oversee governance¹⁹⁹. Whilst learning anatomy using cadavers has a long history as outlined above, the legally permitted use of human cadavers explicitly for the purposes of training surgeons is therefore a surprisingly recent phenomenon – it has only been legal for the last 15 or so years.

2.5.3 Ethical considerations

I think it is important to consider the ethical challenges with using human bodies for training surgeons. There is already a well-developed body of opinion as to the proper ethical treatment of cadavers in the undergraduate anatomy setting²⁰⁰⁻²⁰². The issue

of using cadavers for training surgeons raises different issues and warrants separate consideration.

Unlike purely anatomic dissection, where the body tissues are explored largely unaltered from their natural state, performing surgery (simulated or otherwise) by definition necessitates the removal, rearrangement and sometimes destruction or damaging of body parts or tissues. A simulated surgical procedure may also involve implanting foreign materials into the cadaver, in the case of T&O that might involve deliberately breaking a bone to enable a trainee surgeon to practice fracture fixation with metal plates and screws.

This distinction in purpose and practice from anatomy dissection has implications for informed consent for body donation. It also raises questions about the extent to which we are using the body as an object to be manipulated, and even destroyed.

There is also the issue of commercialisation to consider. Cadaveric surgical training courses, in contrast to undergraduate anatomy training within medical schools, rely heavily on for-profit donation companies to meet the demand for cadaveric material²⁰³. I used such a company to provide the cadavers for the trial detailed in chapter 5 of this thesis, as we do not (yet) have a regional body donation programme. This raises questions of turning deceased humans into commodities that can be bought and sold, and whether this might influence the motives of prospective donors and their families in considering a bequest.

Given the apparently rapid expansion in provision of cadaveric simulation training, it is important to consider these issues. How we as a surgical profession treat deceased bodies reflects societal values and attitudes towards the dead, and the respect necessarily shown to them because of their human status.

2.5.4 Is high fidelity always best?

In designing a simulation training intervention, there is an important question to be asked about fidelity. A simplistic definition of fidelity is 'realism', and as outlined

above, it has been variously described as having physical, environmental, equipment and psychological subcategories.

Fidelity is relevant here for two reasons. Firstly – fidelity appears to be related to training effectiveness, where effectiveness is defined as transfer of skills. This relationship has been described as the ‘fidelity-transfer correlation’²⁰⁴.

The conventional view is that the level of fidelity of simulation should match the skill level of the learner and the educational goal of the training exercise¹³⁵, in other words, more expert level learners require higher fidelity simulation. The second reason why fidelity is important is cost – as a general rule, high fidelity simulation is more expensive than low fidelity simulation²⁰⁵.

In their classical work on the fidelity-transfer correlation of training systems, Miller²⁰⁶ formalized this apparent relationship between fidelity, degree of transfer, and cost. Importantly, Miller identified a point of diminishing returns, at which point costs escalate without associated increase in training benefit (figure 9).

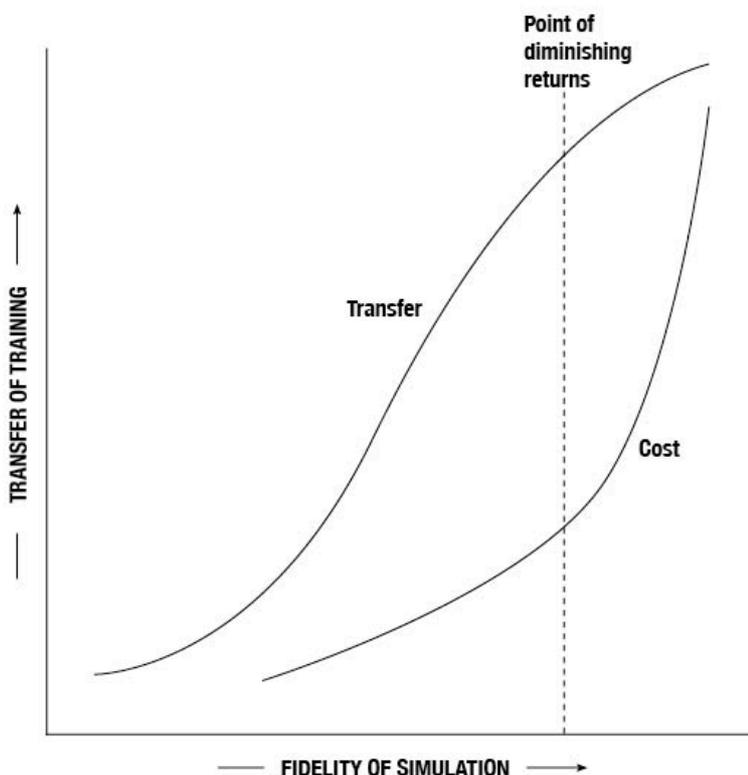


Figure 9. Miller’s putative relationship between fidelity, cost and training effectiveness²⁰⁶

A criticism of this approach is that it ignores the different levels of experience of learners, so Alessi and colleagues²⁰⁷ augmented Millers' curve by distinguishing learners by expertise. Further research work has continued in the field of aviation simulator research, by adding more contextual variables to the design of simulators to investigate what became known as 'the fidelity question' described by Hays and Singer²⁰⁴; how similar to the actual task must a simulator be to provide effective training?

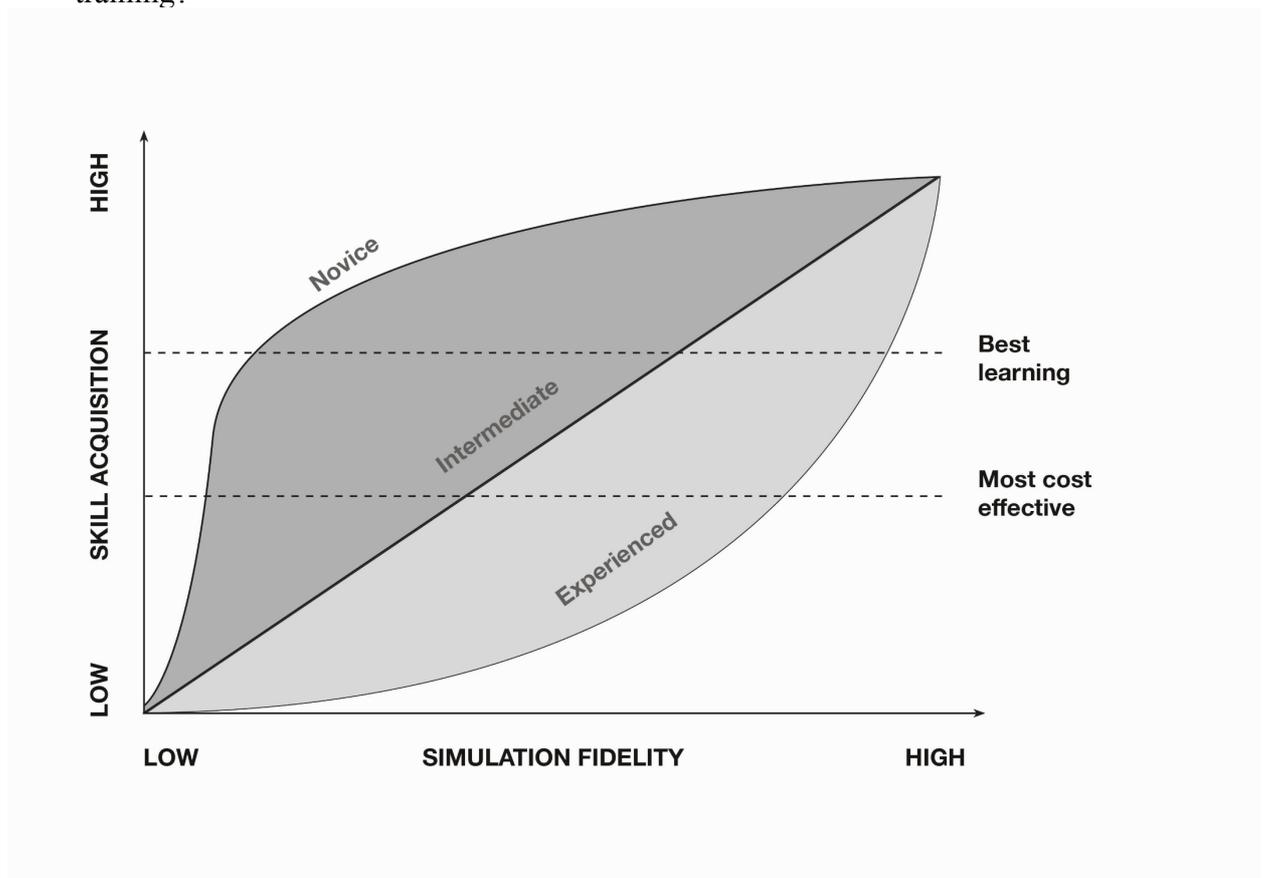


Figure 10. Alessi modifications to the fidelity-transfer correlation (adapted from Aggarwal et al¹³⁵)

Although this work on theory development and the fidelity question comes from the aviation simulation literature, the implications are directly relevant to surgical simulation. The evidence base on optimal simulation fidelity is in its relative infancy, but research evidence that both supports²⁰⁸ and challenges²⁰⁹⁻²¹¹ Millers traditional theory can already be found in the surgical education literature.

I think the overall message here is that the mechanism by which fidelity affects learning is not well understood, and that the ‘optimum’ simulation fidelity for surgery is likely to be highly context specific – dependent on the learner population, the procedure being taught, and the learning gains expected.

This should all be borne in mind when discussing cadaveric simulation training. As I have said above, cadaveric simulation is high fidelity (or ultra-high, depending on who you ask). Cadaveric simulation is expensive. Even when there is a local body donation programme (the biggest cost in delivering this training is often commercial purchase of cadaveric material), there is still a requirement for a specialist laboratory facility, with all the resource implications and costs that brings.

The biggest criticism of cadaveric simulation for surgical training is the question around cost-effectiveness. This is particularly relevant in the climate of financial austerity (described in detail in chapter 1), where it has never been more important to demonstrate value in spending public money.

In this thesis I am investigating the effectiveness of ultra-high fidelity simulation training for relatively inexperienced surgical trainees, who are probably best described as advanced beginners according to the Dreyfus model (please see figure 2). This flies in the face of the traditional thinking around the fidelity question, proponents of which might argue that, according to Miller/Alessi et al, inexperienced trainees would get the greatest learning gains from low fidelity simulation models such as plastic bones and bench-top trainers.

I am challenging this position in my thesis, and I believe that, because of the unique training benefits of cadaveric simulation for learning surgery (and the particular present challenges of the surgical training climate), the conventional rules about the fidelity-transfer correlation do not apply to surgical simulation, and a new model will be proposed. The evidence to support my position on this is shown in chapters 7 and 8, and I will revisit this again in the conclusion chapter (chapter 9).

2.6 Conclusion

The primary application of simulation in the domain of technical skill training is to provide learners with an opportunity for deliberate practice¹³⁵. They can make mistakes, learn from these, achieve and demonstrate competence in a safe environment away from patients. This, in theory at least, shortens the early part of the surgical learning curve (please see figure 3 above) and could safeguard patients from avoidable harms inflicted by inexperienced surgical trainees.

Cadaveric simulation is an exciting new development in surgical training, which, when conducted within the appropriate ethical framework, offers an unparalleled opportunity to practice operations in their entirety in a highly realistic environment.

There are unanswered questions around the effectiveness of ultra-high fidelity cadaveric training for inexperienced surgical trainees, and implicit questions around transfer validity and cost-effectiveness. I will explore these questions further in this thesis.

2.7 Reflections for chapters 1 and 2

In chapters 1 and 2, I have set the scene for the research presented in this thesis. I gave a lot of thought as to how to best explain the current issues within surgical training, the complex landscape within which they are situated, and how to strike the necessary balance between brevity and detail. I felt that a bit of background history around surgical training was helpful to explain how we have got to where we are today, and to illustrate the fact that the Improving Surgical Training pilot feels simultaneously exciting, and yet also a bit like reinventing the wheel (and to the harshest critics, MMC version 2.0).

I hope the reader agrees that the depth of description I have provided on the current challenges to surgical training in chapter 1 is justified. It would have been very easy to say in one short paragraph ‘there is evidence that trainee morale is declining’ and ‘quality of training has been negatively impacted by working hours restrictions and financial austerity in the NHS’, but this would not be doing justice to the scale or complexity of the problem.

I happily accept criticism that the section in chapter 1 on the quality and morale problems in training would appear, in the words of one of my colleagues, to look ‘like a bit of a rant’. It is, I think, a reflection of my passion for the serious issues in training that it comes across this way. I have, clearly, an academic interest in finding solutions to the problems in training, but as a current T&O surgeon-in-training I also have a personal investment in it.

I was a CT2 trainee (postgraduate year 4) who was stuck on the wards unable to get to the operating theatre. I spent almost all my time in CT2 doing paperwork, basic non-surgical clinical tasks on the wards, and getting impatient with the woeful inefficiency of the hospital systems. The demands of the wards were such that the work was never finished, and as there were no house officers or orthogeriatric medical support, the responsibility of day-to-day management of a large cohort of medically complex, unwell inpatients fell to core trainees.

I 'successfully completed' core surgical training (by the metric of the ARCP system at least) with minimal surgical experience. I have never taken out an appendix, never repaired a hernia (and still have no idea how to do either) with the sum total of my operative experience at that stage acquired during rare (interrupted) moments in theatre, before the inevitable call would come to go back to the wards to do some mundane, non-educational task like rewrite a drug chart or replace a cannula. This lack of surgical experience in the 'core' years is the direct result of a system that front-loads training with non-educational service provision which I described in chapter 1.

In addition to the onerous service demands my colleagues and I faced, another major reason for this lack of proper operating experience was that the ST3 registrars, themselves recent products of this same environment, were having to learn the basic operations that they were denied the opportunity to do during core training for the same reasons we were.

When I became an ST3 registrar (woefully surgically under skilled by recent historical standards, as were my colleagues), I finally learnt to do some basic orthopaedic operations, and the CT2 working with me was similarly denied the opportunity as I had been a year earlier. That trainee – who was excellent - subsequently left surgical training and moved into a medical specialty, and cited the poor quality of training in T&O as a contributory reason.

I was fortunate at that time to be in the relatively comfortable position of having a 'run-through' job, by virtue of being on the integrated clinical academic track. Therefore despite the daily realities of my training environment at that stage being totally at odds with our expectations as 'surgical' trainees, I did not have to worry about making my logbook competitive for national selection into ST3. I clearly remember the sense of frustration and anxiety amongst my CT2 colleagues in uncoupled core posts who were not getting to theatre, and whom were inevitably disadvantaged against others from district general hospitals (where the wards were much less busy, patients less sick, and there was often F1 and orthogeriatric medical cover). I also remember very well the common feeling amongst our peer group that

these problems were not being recognized by our consultant trainers, the ward-based work we were doing was not valued, and we were just 'numbers on the rota'.

It was of no surprise therefore to read of the scale of trainee dissatisfaction in the results of the GMC national trainee survey. Clearly what my colleagues and I were experiencing was representative of what was happening all over the country. It was also disappointing, although again unsurprising, to learn that applications to surgical training are falling, that trainees are leaving in ever-increasing numbers, and it is increasingly in danger of no longer being seen as an attractive career option.

I have harboured a passion for orthopaedics since the earliest days of medical school, fostered by some excellent early career mentors. My enthusiasm remains unwavering, but is tempered by a growing sense that trainees are being let down by a system that does not afford them the amount or quality of training they need to be effective 'day 1' consultants. The fact that the current situation has been allowed to evolve is, I believe, a rather damning failure of leadership by our current training and regulatory bodies.

Having said all this, I do believe there is reason to be optimistic about the future of training. There is a growing awareness of the problems that I have described here, and an enthusiastic body of surgical education researchers who are beginning to provide evidence for workable solutions. I am happy to be part of this movement, and to apply the frustration I have felt at the inadequacy of my own early training experiences in a constructive way as motivation towards the delivery of this PhD thesis.

Chapter 3: Cadaveric simulation for postgraduate surgical training: systematic review of the literature

In this chapter I will systematically appraise the evidence base for the current use of cadaveric simulation for surgical training. It is important to understand the existing evidence of educational impact, both to identify gaps, and to situate my trial within the current knowledge base.

Declarations

This chapter has been published;

James HK, Chapman AW, Pattison GTR, Griffin DR, Fisher JD. Systematic review of the current status of cadaveric simulation for surgical training. *British Journal of Surgery*. 2019;106(13):1726-34.

James HK, Chapman AW, Pattison GTR, Griffin DR, Fisher JD. Re: Cadaver Simulation in Surgical Training. *British Journal of Surgery*. 2020;107(4):468.

James HK, Pattison GTR. Author response to: Comment on: Systematic review of the current status of cadaveric simulation for surgical training. *British Journal of Surgery*. 2020;10.1002/bjs.11629.doi:10.1002/bjs.11629.

Mrs A W Chapman (AWC), consultant orthopaedic surgeon, provided independent data extraction as second reviewer and independent assessment of references for inclusion.

3.1 Introduction

There has been rapid expansion in the provision of CST courses¹⁷⁶ since the HTA rules changed to allow cadaveric training for surgeons¹⁸⁰. To understand the current evidence base for using cadaveric simulation in postgraduate surgical training a systematic review of the literature is required.

The move to incorporate simulation into surgical training is driven by a need to improve training efficiency and safety, as discussed in chapter 1. Of all the available surgical simulation modalities, cadaveric simulation is of particular interest, as for the reasons described in chapter 2, it provides highly realistic representation of surgical anatomy as encountered “*in vivo*”^{7 176} and authentic tissue handling/haptic feedback characteristics¹⁸⁵. It allows for appreciation of complex three-dimensional neurovascular relationships which are difficult to learn from textbooks, and almost impossible to replicate in synthetic low-fidelity models¹⁸³.

Cadaveric simulation also offers surgical trainees the opportunity to practice an operation in its entirety ‘skin-to-skin’ with high environmental, equipment and psychological fidelity, without the time pressures of the real life operating theatre. It enables the rapid acquisition of procedural skills⁷ and attainment of competence in a time and cost-efficient manner. By training junior surgeons-in-training in the cadaveric laboratory, the early part of the surgical learning curve takes place in a setting remote from patient care, which might also safeguard patients from inexperienced surgeons⁸⁴.

With the current enthusiasm for, and increasing provision of cadaveric training courses for surgical trainees²¹², a systematic evaluation of the evidence for their use is timely and necessary. This review is also necessary to identify evidence gaps and to situate the trial in this thesis (chapter 7) within the current evidence base.

3.2 Aim

The aim of this chapter is to describe and appraise the evidence for the use of cadaveric simulation in postgraduate surgical training

3.3 Methods

This systematic review was undertaken in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines^{213 214} and was registered with PROSPERO (International prospective register of systematic reviews) (CRD 42018110426)²¹⁵.

3.3.1 Data Sources

I conducted the literature search in January 2019. The databases searched included; MEDLINE (Ovid) (1946-present), CINAHL (EBSCO) (Cumulative Index of Nursing and Allied Health Literature), Centre for Reviews and Dissemination Database, ISRCTN Registry, Cochrane Central Register of Controlled Trials, NHS Evidence, Pubmed (1950-present), Embase (Ovid) (1947-present), Scopus, Australian Clinical Trials Registry and Google Scholar.

3.3.2 Search strategy

My search included Medical Subject Headings where appropriate and search results were limited to Human Subjects and English Language. The search strategy (table 1) was developed in MEDLINE and adapted according to the required syntax of the different databases.

Table 1: Indicative search strategy

MEDLINE Search strategy*
Category A: 'cadaver' OR 'cadaver*.mp' Category B: 'Clinical Competence/ OR Simulation Training/OR simulat*.mp' Category C: 'Internship and Residency'/OR Clinical Competence/ OR Curriculum/ or surgical training.mp. OR Education, Medical, Graduate/.
Limits: Human Subjects and English Language. No date limits applied.

*Categories A, B and C were combined with the Boolean operator 'AND', and the specified limits applied.

I removed the duplicates and obviously irrelevant studies. The titles and abstracts were screened for initial eligibility by me and the second reviewer (AWC). The flow of studies in the review is shown in figure 11. Reference lists of the included studies and old reviews were hand-searched.

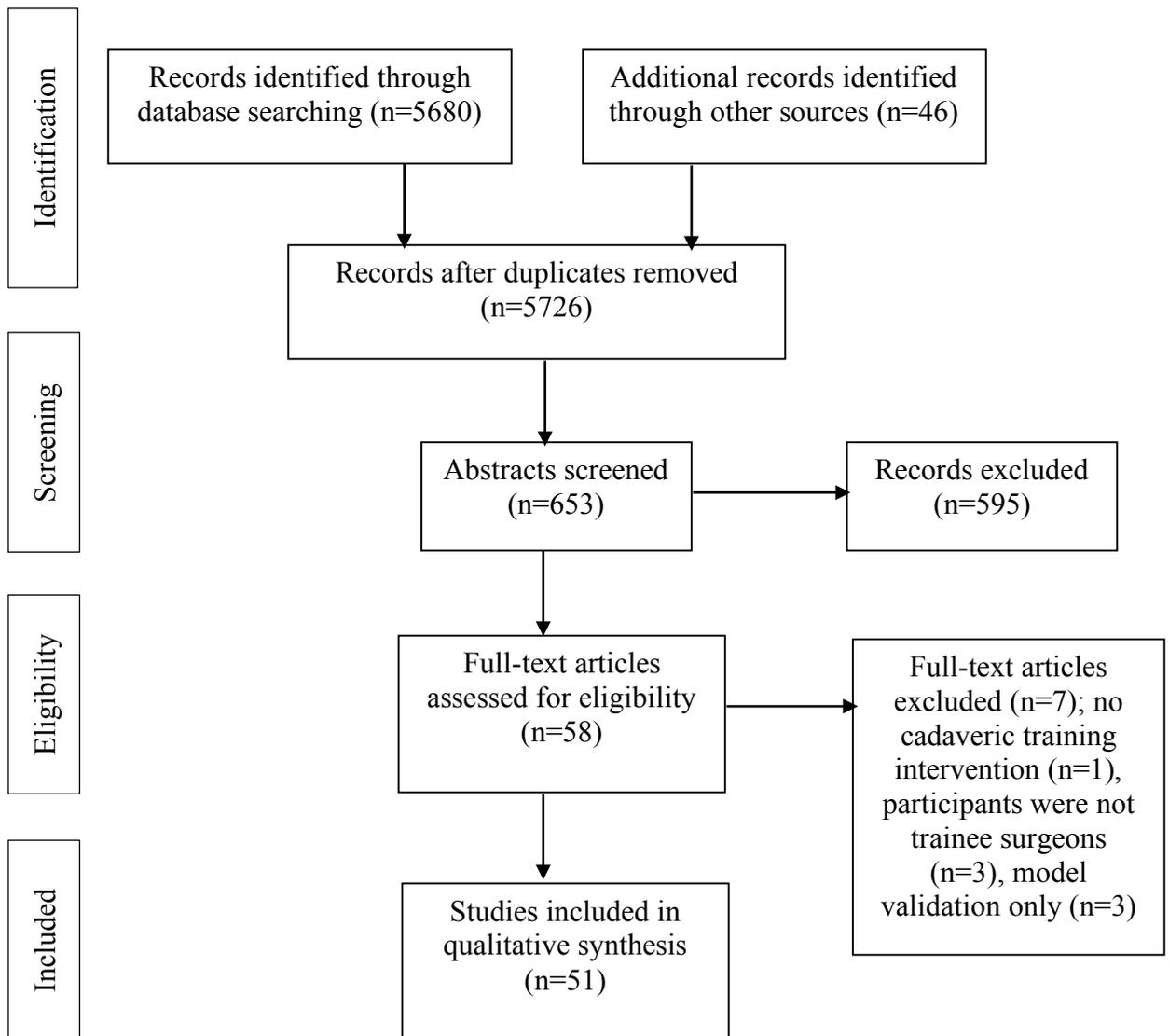


Figure 11. PRISMA Flow-chart of included studies

3.3.3 Selection Criteria

Inclusion criteria;

- 1) Study participants were postgraduate doctors-in-training
- 2) There was a human cadaveric simulation training exposure
- 3) There was an attempt to measure the educational impact of the training
- 4) Full-text available in English Language.

Exclusion criteria;

Studies were excluded at screening if they used animal cadaveric models, the participants were veterinary trainees or medical undergraduates. Studies that just described a cadaveric simulation technique without any attempt at assessing the educational impact were also excluded.

3.3.4 Data Extraction

Abstracts that passed eligibility screening were retrieved in full-text (n=58). Reference lists of full-text articles were examined for relevant studies. Additional material found by hand searching were subject to the same eligibility screening process.

The data was extracted from the full text articles using piloted data extraction forms, with myself and AWC working independently (please see Appendix 11.1 for data extraction form).

Data items collected included;

- 1) **Participant characteristics;** number, stage of training and surgical specialty
- 2) **Study design;** single vs multi-centre, eligibility criteria, loss to follow-up with reasons
- 3) **Cadaveric Training;** intervention, cadaveric model used, skills taught, comparator group (where applicable)
- 4) **Assessment of educational impact;** primary outcome measure, evidence of instrument validation, results summary (objective and subjective), post-test assessment and evidence of skill transfer (where applicable).

3.3.5 Quality Assessment

I assessed the quality of evidence by assigning each study a Level of Evidence Score (LoE) using a modified version of the Oxford Centre for Evidence Based Medicine (OCEBM) classification. This has been adapted by the European Association of Endoscopic Surgery and has been widely used in educational systematic reviews¹⁴¹²¹⁶. I assessed the methodological rigor of the reviewed studies using the Medical Education Research Quality Instrument (MERSQI), which is a previously validated assessment tool²¹⁷⁻²¹⁹ for quantitatively appraising medical education research. MERSQI appraises studies across six domains; sampling, type of data, instrument validity, data analysis and outcome. The maximum possible score is 18 points. Both AWC and I assigned each included study a MERSQI score, and the mean score of both reviewers is presented in the results.

Due to the heterogeneous nature of the included studies, I decided to do a qualitative, narrative synthesis of the evidence. This was structured around an adapted Kirkpatrick's hierarchy (figure 8, chapter 2), which is an accepted framework for assessing the educational impact of a training intervention.

3.4 Results

My initial literature search generated 5,726 unique citations, of which 5,073 were clearly ineligible and rejected at title review. 653 abstracts were screened by both myself and AWC, 595 did not pass eligibility screening and were excluded. 58 articles were accessed in full-text and carefully reviewed, and seven studies were rejected at this stage. Three studies were rejected because the participants were consultants and not trainee surgeons, three were rejected because the studies concerned simulator model validation with no attempt at educational assessment, and one study was rejected at this stage it was subsequently found to have no cadaveric

simulation training intervention, although this was not clear from the abstract. 51 articles were included in the review (shown in summary form in tables 4a-d).

Of the 51 studies, four were conference posters and 47 were full-text original research articles.

3.4.1 Study design and setting

Eight of the 51 studies included in the review were randomised controlled trials, six were comparative cohort studies, and 37 were non-comparative cohort studies (tables 4a-d).

The majority of the studies were from the USA (35 studies) and the UK (eight studies), with the remainder from Canada (four studies), Australia (two studies) and one each from Germany and Japan. All the included studies except for one²²⁰ were single-centre for the delivery of the educational intervention.

3.4.2 Study participants

The number of study participants in the review studies varied widely from three to 390, totalling 2002 individual trainee surgeon participants across 69 separate cadaveric training interventions.

3.4.3 Surgical specialty

A wide range of surgical specialities were represented in the review studies. The majority were within general surgery (14 studies), T&O (nine studies) and neurosurgery (seven studies). All of the included studies were single specialty. Study grouping with mean MERSQI scores are shown below in table 2.

Table 2. Included studies grouped by surgical specialty

Surgical Specialty	Number of studies	Mean MERSQI score
General Surgery	14	9.6
T&O	9	11.2
Neurosurgery	7	8.2
Vascular Surgery	5	9.4
ENT	3	12.5
Plastic Surgery	3	7.5
Urology	3	6.5
Obstetrics & Gynaecology	2	10.3
Dermatology	2	8.3
Ophthalmology	1	10
Emergency Medicine	1	7
Maxillofacial Surgery	1	5.5
Total	51	9.4

3.4.4 Study quality

The mean MERSQI score was 9.4 (range 3.5-14) (Table 3). In terms of level of evidence, only two of 51 studies were OCEBM level 1b (RCT of good quality and adequate sample size with a power calculation), six studies were OCEBM 2a (RCT of reasonable quality and/or of inadequate sample size), six studies were OCEBM 2b (parallel cohort study) and 37 studies were OCEBM level 3 (non-randomised, non-comparative trials, descriptive research).

A linear relationship is observed between Kirkpatrick level and mean MERSQI score, suggesting that quality of evidence is positively associated with robust methodology. Table 3 shows an overview of the level of evidence and MERSI scores by Kirkpatrick level.

Table 3. Studies by Kirkpatrick level; OCEBM classification and MERSQI scores (range and mean).

Kirkpatrick Level	Number of studies	OCEBM level of evidence					MERSQI score	
		1b	2a	2b	2c	3	Mean	Range
1 - Reaction	22	-	-	1	-	21	6.4	3.5-10
2 - Knowledge	5	-	1	1	-	3	9.5	7.5-12
3 - Behaviour	23	2	5	5	-	11	12.3	8-14
4 - Results	1	-	-	-	-	1	13.5	-
Total	51	2	6	7	0	36	9.4	3.5-14

3.4.5 Assessment of educational impact

28 of 51 studies assessed the educational impact of the training intervention using objective measures. 23 of 51 studies used subjective measures only to assess the educational impact. Of the 28 studies that used objective outcome measures, 16 of these attempted to measure skill transfer following training. The study characteristics, methods and results are summarized in tables 4a-d.

Table 4a: Kirkpatrick Level 1: Studies that subjectively measure the impact of the training intervention by learner opinion							
Study	Training Intervention	Skills Taught	Comparator	Primary Outcome Measure	Results	Skill Transfer	MERSQI Score
OCEBM Level 2b: Comparative cohort design							
<i>Gunst et al.</i> ²²¹ , 2009 Trauma Surgery N=18 PGY NS	8 x 1 day course	Trauma surgical exposures	n/a	Operative confidence	Increase in confidence scores for 44/48 exposures (p<0.0001), no decline at 6/12	N	7.5
<i>Sharma et al.</i> ²²² , 2012 General Surgery N=45 PGY 1-5	Supervised performance of tasks on FFC	Level specific; senior = lap sigmoid colectomy, intermediate = lap incisional hernia repair, junior = basic laparoscopic tasks	VR trainer	Fidelity	CST perceived as significantly better model overall vs high fidelity VR by all grades	N	6
<i>Gasco et al.</i> ²²³ , 2013 Neurosurgery N=6 PGY NS	57 cadaveric dissections within multi-modality simulation curriculum	Curriculum based	n/a	Learner opinion	CST highest benefit vs other modalities (p<0.001). Effect was greater with junior vs senior	N	10

OCEBM Level 3: Non-randomised, non-comparative descriptive studies							
<i>Dunnington et al.</i> ²²⁴ , 2003 General Surgery N=9 PGY 2-3	2 hour skill session	Sentinel node mapping and excision, level I/level II axillary dissection	n/a	Fidelity, operative confidence, learner opinion	Valuable demonstration of procedure, practice opportunity and confidence	Y	6
<i>Supe et al.</i> ²²⁵ , 2005 General Surgery N=32 PGY NS	Laparoscopy training course with lab	Cholecystectomy, splenectomy, appendectomy, mesenteric lymph node biopsy	n/a	Learner opinion	31/32 reported high satisfaction with training model, 29/32 anatomical learning	N	6
<i>Giger et al.</i> ¹⁷⁵ , 2008 General Surgery N=33 PGY NS	6 x 1 or 2 day courses taught as 3 modules over 2 years	Advanced laparoscopic skills; colon/hernia/bariatric/vascular	n/a	Fidelity	Authenticity of tissue colour, tissue consistency and operative tactility	N	12.5
<i>Reed et al.</i> ²²⁶ , 2009 Vascular Surgery N = 45 PGY 2-3	6 discrete sessions over 24 months	Carotid endarterectomy, fem- pop bypass, SMA embolectomy, 4 compartment fasciotomy	n/a	Learner opinion	100% optimal training material, 97.8% rated the educational value of the course as perfect	N	6
<i>Lewis et al.</i> ²²⁷ , 2012 General Surgery N = 150 PGY 1-5	Weekly educational programme	Not specified	n/a	Leaner opinion	Positive opinion of cadaveric sessions, learning stages of operations and increased confidence	N	6

<i>Shekter et al.</i> ²²⁸ , 2013 Plastic Surgery N = 192 PGY 1-6	Weekly afternoon cadaveric lab	Curriculum based	n/a	Operative confidence	Overall confidence was improved, 1.90 to 4.20 (p<0.001)	N	7
<i>Jansen et al.</i> ²²⁹ ,2014 Vascular Surgery N = 26 PGY NS	2 day cadaveric lab	Open vascular exposures	n/a	Learner opinion	100% agreement with course meeting learner objectives	N	6
<i>Pham et al.</i> ¹⁸⁴ , 2014 Neurosurgery N = NS PGY 3-5	Single procedural attempt on cadaver	Management of intraoperative ICA injury	n/a	Operative confidence and procedural knowledge	Improved knowledge and procedural confidence	N	6.5
<i>Aboud et al.</i> ²³⁰ , 2015 Neurosurgery N = 203 PGY NS	13 cadaveric courses over 5 years	Management of intracranial aneurysm rupture	n/a	Fidelity, learner opinion	Positive for fidelity, skill acquisition, valuable to training	N	5
<i>Ahmed et al.</i> ²³¹ , 2015 Urology N = 81 PGY 3-10	3 day modular cadaveric course	Curriculum based	n/a	Fidelity, operative confidence, learner opinion	All procedures scored a mean of 3/5 for face validity. Subjective improvement in skills and transferrable skills for the operating room	N	6

<i>Aydin et al.</i> ²³² , 2015 Urology N = 42 PGY NS	Urology Human Cadaver Training Programme	Not specified	n/a	Fidelity, operative confidence, learner opinion	Mean = 3/5 for face validity. Useful for learning anatomy, confidence boosted, skills improved and feasible for training	N	6
<i>Liu et al.</i> ²³³ , 2015 Neurosurgery N = 12 PGY 2-7	4 or 5 x 4 hour sessions	Cranial and endonasal approaches, temporal drilling technique	n/a	Learner opinion	Improved anatomy and liked rehearsal opportunity. 67% applied learning to real life operating	Y	5
<i>Winer et al.</i> ²³⁴ , 2015 Neurosurgery N = 203 PGY NS	Feasibility study	Neuroendoscopic procedures	n/a	Operative confidence, learner opinion	Improved knowledge, confidence	N	5
<i>Clark et al.</i> ²³⁵ , 2016 Dermatologic surgery N = 8 PGY 2-4	Cadaveric lab practicum	Digital blocks, nail plate avulsions and nail matrix biopsies	n/a	Operative confidence and procedural competency	Improvement in confidence and competency (post vs. pre) (p<0.0001 for both)	N	7
<i>Hanu-Cernat et al.</i> ²³⁶ , 2016 Maxillofacial surgery N = 41 PGY 3-7	Cadaver course	Not stated	n/a	Operative confidence	Senior participants confidence improved by 3 points, junior by 4 points	N	6.5

<i>Calio et al.</i> ²³⁷ , 2017 Neurosurgery N = 17 PGY 1-5	Not specified	Spinal decompression and fusions	n/a	Learner opinion	Cadaveric training superior to other simulation modalities (94%)	N	6
<i>Lorenz et al.</i> ²²⁰ , 2017 General Surgery N = 390 PGY NS	0.5 day cadaveric lab within 3 day course	Open and endoscopic hernia repair	n/a	Learner opinion	95% stated improved training	N	6
<i>Nobuoka et al.</i> ²³⁸ , 2017 Hepatobiliary Surgery N = NS PGY NS	Thiel cadaver workshop	Open + laparoscopic Whipple's and major hepatectomy	n/a	Learner's opinion	Excellent satisfaction and good understanding of surgical anatomy was achieved in many participants	N	3.5
<i>Pacca et al.</i> ²³⁹ , 2017 Neurosurgery N = NS PGY NS	'hands-on' dissection course	Management of sharp and blunt ICA injury	n/a	Learner Opinions	100% reported gaining new knowledge, learning new techniques and interest in further training on the model	N	6
<i>Takayesu et al.</i> ²⁴⁰ , 2017 Emergency Medicine N = 22 PGY 3-4	1 hour cad session after mannequin teaching	Surgical cricothyroidotomy and tube thoracostomy	Low fidelity simulator and animal chest walls	Simulation fidelity and operative confidence	Higher fidelity of cadaveric training vs mannequins (p<0.0001) for both procedures	N	7

<i>Weber et al.</i> ²⁴¹ , 2017 Plastic Surgery N = 9 PGY 1-7	Single procedural attempt within 7/7 of live attempt	Not stated	n/a	Operative confidence	Improved median confidence score (pre-simulation 2/5 and post simulation 4/5) p<0.01	N	7
<i>Bertolo et al.</i> ²⁴² , 2018 Urology N = 22 PGY 4-7	1 day intensive robotic course	Pelvic and kidney robotic procedures via the transperitoneal approach	n/a	Operative Confidence, Learner Opinion	Improved confidence in 5 robotic skill domains	N	6.5
<i>Chouari et al.</i> ²⁴³ , 2018 Plastic Surgery N = 50 PGY 1-7	Annual 2 day course, first day of 2 courses evaluated	Microsurgical flap reconstruction	n/a	Operative Confidence	Improved confidence (p<0.005) and preparedness for unsupervised bench work	N	7.5
<i>Wynn et al.</i> ²⁴⁴ , 2018 General Surgery N = NS. Stratified by years of experience; <5yrs, 5-10yrs or >10yrs.	Specialist cadaveric training course	Transanal total mesorectal excision	n/a	Operative exposure post-training	47% had performed between 1 and 13 cases within 6/12	Y	8.5

Table 4b: Kirkpatrick Level 2: Studies that objectively measure the impact of the training intervention by learner knowledge							
Study	Training Intervention	Skills Taught	Comparator	Primary Outcome Measure	Results	Skill Transfer	MERSQI Score
OCEBM Level 2a: Randomised Controlled Trial							
<i>Al Jamal et al.</i> ²⁴⁵ , 2018 General Surgery N = 14 PGY 1-4	Single supervised performance on a cadaver	Endoscopic total inguinal hernia repair	Low fidelity bench top simulator	Procedural knowledge scores	No difference seen in test scores between two modalities	N	10.5
OCEBM Level 2b: Comparative cohort studies							
<i>Sharma et al.</i> ²⁴⁶ , 2016 General Surgery N = 14 PGY 2-3	8 x 2 hour cadaveric lab sessions across 8 weeks	Open cervical, thoracic, abdominopelvic and extremity procedures	Course materials only	Viva Voce examination scores	Larger improvement in overall examination scores in cadaveric group (31% +/- 4% vs. 8% +/- 3%, p=0.0006)	N	9.5
OCEBM Level 3: Non-randomised, non-comparative descriptive studies							
<i>Mitchell et al.</i> ²⁴⁷ , 2012 Vascular Surgery N = 22 PGY 3-4	5 X 4 hour cadaveric lab sessions	Complex open vascular surgical approaches	n/a	Oral checklist exam	Mean examination scores significantly improved across all 5 exposures	N	12
<i>Robinson et al.</i> ²⁴⁸ , 2017 Vascular Surgery N = 58 PGY 1-7	0.5 days cad lab as part of course	Open surgical approaches	n/a	Procedural knowledge scores	Improvement in knowledge scores in both groups	N	7.5

<i>Hazan et al.</i> ²⁴⁹ , 2018 Dermatologic Surgery N = 40 PGY 1-4	Cadaver training session	Not stated	n/a	Procedural knowledge scores	Overall improvement in pre and post intervention knowledge (p=0.001)	N	8
---	-----------------------------	------------	-----	-----------------------------------	--	---	---

Table 4c: Kirkpatrick Level 3: Studies that objectively measure the impact of the training intervention by change in learner behaviour							
Study	Training Intervention	Skills Taught	Comparator	Primary Outcome Measure	Results	Skill Transfer	MERSQI Score
Randomised Controlled Trials; Cadaveric Simulation vs. No Simulation							
<i>Sharma et al.</i> ²⁵⁰ , 2013 General Surgery N = 19 PGY 1-2	10 repetitions of 5 tasks across 2 days	Peg transfer, simulated appendectomy, intra and extra corporeal knot tying	No training	GOALS scale performance on cadavers, pre-/post intervention, VR performance	4/5 tasks on cadavers showed significant improvement on learning curve analysis. Post-test VR assessment showed safety of cautery (p=0.4) and left arm path (p=0.047) to be significantly improved in the intervention group	Y	14
<i>Sundar et al.</i> ²⁵¹ , 2016 Neurosurgery N = 8 PGY 1-4 and 2 medical students	Cadaveric training course	Pedicle and lateral mass screw placement	Didactic teaching	Final Product Analysis, surgical error	Reduced surgical error in the cadaveric vs. control group (p=0.04). Screw placement was more optimal in cadaveric vs control group in cervical,	N	14

					thoracic and lumbar regions (p=0.02, 0.04 and 0.04)		
<i>Chong et al.</i> ²⁵² , 2017 O&G N = 34 PGY 1-4	0.5 day cadaveric lab	Transobturator tape insertion	Didactic teaching	Procedure scores on low fidelity mannequin	Procedure scores improved in intervention group vs. controls p<0.01	Y	11
Randomised Controlled Trials; Cadaveric Simulation vs. Low-fidelity Simulation							
<i>Anastakis et al.</i> ²⁵³ , 2003 General Surgery N = 23 PGY 1	4 hour cadaveric lab session	Basic general surgical skills	Low fidelity simulator and text materials only	Procedure checklist score and GRS	Significant effect of training modality on checklist and GRS scores. Bench and cadaveric training were superior to text, and bench and cadaveric training were equivalent to each other	Y	14
<i>Sidhu et al.</i> ²⁵⁴ , 2007 Vascular Surgery N = 27 PGY 1-4+	3 hour cadaveric lab session	Graft-to-arterial anastomosis	Benchtop simulator	Procedural checklist, GRS, final product analysis	Juniors practicing on cadaveric model performed better on checklist (p=0.05) and final product analysis (p=0.04). Seniors practising on cadaveric	Y	14

					models scored better on final product analysis		
<i>Gottschalk et al.</i> ²⁵⁵ , 2015 T&O Surgery N = 15 PGY 1-6	Cadaveric workshop	Cervical lateral mass screw placement	1. Sawbones workshop 2. No training	Final Product Analysis	Both sawbones and cadaveric trained groups improved vs. no training ($p < 0.0001$), sawbone group had modestly higher improvement post intervention than cadaveric group (mean aggregate difference from perfect screw placement -8.2 degrees and -7.2 degrees)	N	14
<i>Camp et al.</i> ²⁵⁶ , 2016 T&O surgery N = 45 PGY 1-5	4 hour cadaveric lab	Knee arthroscopy	VR simulator and no training	Procedure time and checklist scores	Cadaveric trained group improved ASSET scores by 1.1 points per hour of training vs. 0.5 for VR group. Significant decrease in operating time seen in cadaveric group pre- vs post training ($p = 0.002$)	Y	14

OCEBM 2b: Parallel Cohort Studies; Cadaveric vs. Low-fidelity Simulation							
<i>LeBlanc et al.</i> ²⁵⁷ , 2010 General Surgery N = 28 PGY NS	Single performance on cadaver as part of 1 day course	Laparoscopic sigmoid colectomy	VR	OSATS assessment during training	Technical skills scores overall better on VR simulator than cadaver	N	13
OCEBM 2b: Parallel Cohort Studies; Inexperienced vs. Experienced performance							
<i>Zirkle et al.</i> ²⁵⁸ , 2007 ENT N = 19 PGY 1-4	Performance under exam conditions	Cortical Mastoidectomy	Experienced performance	GRS, Final Product Analysis, Task Based Checklist	GRS, TBC and EO correlated with trainee experience. FPA did not. TBC correlated with EO	N	13
<i>Mackenzie et al.</i> ²⁵⁹ , 2017 Trauma Surgery N = 40 PGY3-6	ASSET Course	Lower extremity vascular exposure, repair and fasciotomy	Pre, post (immediate <4/52), post (delayed 12-18mths) vs experts	Procedure scores, GRS, error, frequency, procedure time	Decreased errors from 60 to 19% after training, improved knowledge and procedural steps (p<0.001). Interval experience rather than time since training	Y	14

					affected skill retention up to 18 months later		
<i>Mednick et al.</i> ²⁶⁰ , 2017 Ophthalmology N = 11 PGY 2	Single procedural attempt	Corneal rust ring removal	Experienced performance	Time error rate, final product analysis	Procedure time longer for inexperienced group (187 vs 117 secs mean), rust removal percentage similar (61 vs 69%), not statistically significant	N	10
OCEBM 2b: Parallel Cohort Studies; Within subject performance comparison							
<i>Wong et al.</i> ²⁶¹ , 2004 General Surgery N = 9 PGY 2-4	Cadaveric lab practicum with embedded study	Saphenous venous cutdown	n/a	Procedure time, final product analysis	Decreased mean incision size 4.6 vs 3.4cm (p<0.05), mean time taken to completion of procedure 360 vs 232 seconds (p<0.05), % subjects experiencing complications 37.5% vs 0% (<0.05)	N	9.5
OCEBM 3: Non-comparative studies, descriptive research							
<i>Rowland et al.</i> ²⁶² , 1994 T&O Surgery N = 12 PGY NS	Supervised task performance	Endoscopic carpal tunnel release	n/a	Final product analysis	38% incomplete release of transverse carpal ligament, 17% showed complications	N	8

<i>Levine et al.</i> ²⁶³ , 2006 O&G N = 29 PGY 2-3	5 half day cad sessions	Gynae laparoscopy, salpingectomy, salpingostomy	n/a	Operational metrics	Decrease in mean bead transfer time of 38.5 seconds (p=0.02), median increase in number of beads transferred 2.5 (p=0.0001), median decrease in suture time was 63.5 seconds (p=0.001)	Y	9.5
<i>Bergeson et al.</i> ²⁶⁴ , 2008 T&O surgery N = 3 PGY 1+3	Serial performance on 5 spines	Thoracic pedicle screw placement	n/a	Final product analysis	Decrease in the proportion of total screw violations by the third cadaver (p<0.001) and in critical screw violations by the fourth cadaver (p=0.01)	Y	9.5
<i>Tortolani et al.</i> ²⁶⁵ , 2013 T&O Surgery N = 3 PGY NS	2 cadaveric lab sessions	Thoracic pedicle screw placement	n/a	Final product analysis	Placement accuracy improved following training and structured feedback (44% accurate placement rate pre- and 58% post p=0.01)	N	9.5
<i>Mowry et al.</i> ²⁶⁶ , 2014 ENT	Weekly course over 9 months with self-directed	Microdissection of temporal bone	n/a	PBA, GRS and	Test scores from most recent resident year correlated well with	N	12.5

N = 56 PGY 1-4	access to cadaveric lab			performance time	number of CTB's drilled during corresponding year (r=0.42, p=0.002) and strong correlation between score during the highest year of training and cumulative number of CTB's drilled during residency (r=0.604, p=0.005)		
<i>Awad et al.</i> ²⁶⁷ , 2015 ENT N = 32 PGY 2-6	2 x 3 hour sessions	Mastoidectomy	n/a	Task specific	Longitudinal assessment showed significant improvement with iteration	Y	12.5
<i>Egle et al.</i> ²⁶⁸ , 2015 General Surgery N = 14 PGY 1-5	Cadaveric lab based anastomosis workshop, then repeated 1 week later	Vascular and hand-sewn bowel anastomosis	n/a	OSATS, procedure time and final product analysis	1) Vascular anastomosis: operating time decreased, OSATS scores improved and final product analysis improved (anastomosis leak pressure 38.9 vs 71.8psi p=0.001) 2) bowel anastomosis: operating time decreased	Y	12.5

					(23 vs 18 minutes p<0.001), OSATS scores improved (12.9 vs 14.4 p<0.001), leak pressures improved (17.7 vs 26.9psi p<0.001)		
<i>Nicandri et al.</i> ²⁶⁹ , 2015 T&O Surgery N = 46 PGY NS	3 Day AAOS fundamentals of arthroscopy course	Knee Arthroscopy	n/a	Operational metrics, Checklist scores	9.2% improvement in mean ASSET score (p=0.001), biggest gain seen in those with less than 20 arthroscopic cases (13.2% improvement).	Y	12.5
<i>Kim et al.</i> ²⁷⁰ , 2016 General Surgery N = 133 PGY 1-5	4-6 Cadaveric lab sessions with practice opportunities	General surgical core procedures, stratified by stage of training	n/a	GRS, modified OSATS	% skills that could be performed independently increased from 40-60% (p<0.04)	Y	9.5
<i>Martin et al.</i> ²⁷¹ , 2016 T&O Surgery N = 48 PGY 1-5	4 day course with didactics, low fidelity simulation and cadaveric lab	Knee and shoulder arthroscopy	n/a	Operational metrics using VR simulator pre-and post-test	Improvement post-training in all domains; tip probe distance, time to completion (p<0.001)	Y	12

<i>Ciporen et al.</i> ²⁷² , 2017 Neurosurgery N = 10, PGY NS	3 attempts at procedure under supervision	Endoscopic management of iatrogenic carotid artery injury	n/a	Operational metrics, final product analysis	Time to control bleeding improved with repeated exposure. Senior residents performed better across all performance score domains	N	8
--	---	--	-----	--	---	---	---

Table 4d: Kirkpatrick Level 4: Studies that objectively measure the impact of the training intervention by change in patient outcome

Study	Training Intervention	Skills Taught	Comparator	Primary Outcome Measure	Results	Skill Transfer	MERSQI Score
<i>Martin et al.</i> ²⁷³ , 1998 General Surgery N = 8 PGY 1	Cadaveric lab session	Chest tube insertion, endotracheal tube insertion, venous cutdown	n/a	Correct performance of skill to completion without complication in 2 minutes	All residents passed the 3 skills immediately after instruction. At 3 weeks 8/8 passed VC and CTI, 6/8 passed ETI. Changes in performance over time for both intervals were statistically significant. Times improved after immediate instruction and at 3 weeks (except 2 residents (one each in CTI and VC) and at more than 3 weeks (except one resident in VC) Change in time for both intervals was statistically significant (p<0.05). Complications for all skills decreased significantly immediately and at 3 weeks after instruction (p<0.02)	Y	13.5

ASSET = Arthroscopic Surgical Skill Evaluation Tool
 CST = Cadaveric Simulation Training
 CTB = Cadaveric Temporal Bone
 CTI = Chest Tube Insertion

ETI = Endotracheal Tube Insertion
 EO = Expert Opinion
 FFC = Fresh Frozen Cadaver
 FPA = Final Product Analysis

GRS = Global Rating Scale
 NS = Not stated
 OSATS = Objective structured Assessment of Technical Skill
 PSI = Pounds per Square Inch

TBC = Task Based Checklist
 VC = Venous Cutdown

The studies are grouped for analysis according to how the impact of the educational intervention was measured with respect for the Kirkpatrick hierarchy for learner evaluation¹⁹¹.

Level 1: Reaction – studies that subjectively measured educational impact by learner opinion

Twenty-two of the 51 review studies measured the educational impact of a cadaveric training intervention using subjective measures of learner reaction and/or opinion. One of these studies used a comparative cohort design which compared cadaveric trained with virtual reality trained participants²²² (OCEBM level 2b). Twenty-one were purely descriptive research studies which used non-randomised, non-comparative methods^{184 220 221 223 224 226-229 231-233 235-243} (OCEBM level 3). Two of these Kirkpatrick level 1 studies measured subjective readiness and confidence for real-world procedural performance following the cadaveric training intervention^{224 233}. All of the 22 level 1 studies used pre and or post-training questionnaires to assess learner reaction, most of which were purpose-designed for the individual studies in question²⁴² and not formally validated. The subjective outcome measures collected in the level 1 studies were diverse and included; learner reaction with respect to simulation fidelity, opinion on the usefulness of the training and change in operative confidence and self-perceived competency following the training. All of the level 1 studies reported a positive qualitative effect of the cadaveric simulation training as measured by learner reaction/opinion (table 4a).

Level 2: Learning – studies that objectively measured educational impact by change in learner knowledge

Five studies assessed the educational impact of the cadaveric training intervention by measuring change in learner knowledge. One of these studies was an RCT²⁴⁵ and four were cohort studies^{142 246 248 249}. Three of the level 2 studies used procedural knowledge scores as the primary outcome measure^{245 248 249}, and two studies used traditional *Viva Voce* oral examinations to assess the impact of the training^{246 247}.

AlJamal et al²⁴⁵ ran an RCT comparing a low-cost, low-fidelity benchtop model with cadaveric dissection for teaching endoscopic total extra-peritoneal inguinal hernia repair (TEP-IHR). Fourteen postgraduate year (PGY) 1-4 general surgery residents were randomised to receive training on TEP-IHR on either a cadaveric or low-cost model. Participants underwent a timed, web-based interactive pre-test to assess procedural knowledge. This test was repeated after the training intervention, along with qualitative assessment of the educational experience. The results showed that both groups scored higher on the post-test, with no difference observed between the simulation modality. The survey evaluation results showed that participants preferred the low-cost model over cadaveric training for both initial learning and understanding of the procedure, and participants preferred the cadaveric model in overall experience ($p < 0.05$ for both). No difference was observed in participant preference between the low-cost and cadaveric model in the overall educational value of the training²⁴⁵. The authors concluded that given the marked cost differential between the two models, the 'inexpensive and repeatable inanimate mode' was an 'attractive early resource for learning TEP-IHR'²⁴⁵.

Of the four Kirkpatrick level 2 cohort studies, one²⁴⁶ was a comparative cohort study, comparing learning in cadaveric trained participants with those who received didactic teaching materials only. The study prospectively compared two cohorts of PGY2-3 general surgical residents; the intervention group ($n=7$) received an eight week 'procedurally-oriented' cadaver course, and the control group ($n=7$) received course materials only. Pre- and post- tests of knowledge and operative confidence were administered using a standard-template oral examination and a questionnaire assessing operative confidence. The results showed that overall examination scores were improved in the cadaveric trained group ($31 \pm 4\%$ vs. $8\% \pm 3\%$, $p=0.0006$), this group showed a trend toward higher overall operative confidence, although this was not statistically significant ($p=0.06$).

Three level 2 studies were non-comparative cohorts²⁴⁷⁻²⁴⁹. Significant improvement in post-training knowledge scores was reported in all three of these studies.

Level 3: Behaviour – studies that objectively measured educational impact by learner behavioural change

There were 23 studies that attempted to assess educational impact by measuring change in learner behaviour. The objective behaviour assessment methods that were used across the studies varied widely, and included final product analysis, and various operational metrics (such as motion analysis, error rate, procedure time and instrument/hand path length). There were also several score based methods used, including; OSATS (Objective Structured Assessment of Technical Skills in Surgery), global rating scales (GRS) and GOALS (Global Operative Assessment of Laparoscopic Skills).

Of the 23 Kirkpatrick level 3 studies, 7 were randomized controlled trials (3 comparing CST vs no simulation training²⁵⁰⁻²⁵², and 4 comparing CST vs low-fidelity simulation²⁵³⁻²⁵⁶) and 16 were cohort studies.

Cadaveric vs. no simulator training

Three RCT's compared a cadaveric training intervention with 'no training'. For the purposes of this analysis, provision of didactic learning materials only to the control group has been defined as 'no simulator training'. Chong et al²⁵² compared 17 cadaveric trained (0.5 day course) with 17 untrained (didactic materials only) control participants performance of a trans-obturator tape insertion procedure on a low fidelity simulator, assessed by a blinded expert examiner using a procedural checklist assessment tool. Checklist scores were significantly improved in the cadaveric trained group ($p < 0.01$).

Sharma et al²⁵⁰ compared 10 laparoscopy novices (<3 prior procedures) who had undergone 10 task repetitions (peg transfers, intra/extra-corporeal knot tying, simulated appendectomy) on fresh frozen cadavers with nine controls who had no training. Baseline performance of both groups undertaking a simulated laparoscopic cholecystectomy was measured using the LAPMentor virtual reality simulator.

Cadaveric performances during the training were scored using the GOALS scale (Global Operative Assessment of Laparoscopic Skills Scale) and the VR baseline and transfer performances were assessed by operational metrics. The results showed significant improvement in the safety of intraoperative cautery ($p=0.04$) and left arm path analysis ($p=0.05$). Four of the five cadaveric practice tasks showed significant improvement on learning curve analysis.

Sundar et al²⁵¹ compared five junior neurosurgery residents who underwent a cadaveric training course teaching pedicle and lateral mass screw placement with five stage-matched controls (didactic materials only). Assessment took place one to two weeks after the respective training intervention, participants were each asked to place screws in nine unilateral locations on a cadaveric spine. Assessment was made by computed tomography (CT) scanning the cadaveric specimens and a single blinded assessor scored performance. A significant reduction in ‘overall surgical error’ was seen in the cadaveric group ($p=0.04$) and significant improvement in optimal screw placement was seen in the cervical, thoracic and lumbar regions ($p=0.02$, 0.04 and 0.04 respectively)²⁵¹.

Cadaveric vs. low fidelity simulator training

Four RCT's compared CST with a low fidelity simulation intervention²⁵³⁻²⁵⁶. The results of these studies showed a mixed picture. Sidhu et al²⁵⁴ randomised 27 surgical residents to receive a three-hour ‘hands-on’ training in graft-to-arterial anastomosis on either a low fidelity plastic benchtop model ($n=13$) or high fidelity cadaveric arm ($n=14$). Educational impact was measured by single procedural performance on a live anaesthetized porcine animal model one week post-training, by a single blinded expert assessor using a procedural checklist, global rating scale and final product analysis. Skill acquisition was ‘significantly affected by model fidelity and level of training’²⁵⁴; junior (PGY1-2) participants in the cadaveric trained group performed better than controls on both the checklist ($p=0.05$) and final product analysis ($p=0.04$), and senior (PGY4+) participants in the cadaveric trained group outperformed stage-matched controls on final product analysis ($p<0.05$).

Anastakis et al²⁵³ randomised 23 PGY1 surgical residents into three groups; cadaveric training, bench model training or text materials only, to receive training on six surgical procedures (burr hole insertion, chest tube insertion, small bowel anastomosis, abdominal wound closure, flexor tendon repair and k-wire fixation of fractured metacarpal). Each participant performed each of the six procedures on a cadaver one week after completion of the course. Performance was assessed using checklist and global rating scales by a panel of 12 examiners who rotated between the stations (two examiners per station). Results showed a significant effect of training modality on both checklist and GRS scores, with bench and cadaveric training both superior to text materials only (control group), and that bench and cadaveric training was equivalent. The authors concluded that ‘training on a bench model transfers well to the human model, suggesting potential for transfer to the operating room’.

Camp et al²⁵⁶ randomised 45 PGY1-5 residents into three groups; controls, cadaveric training group and virtual reality simulator group. Baseline and post-test performance was measured in all groups on cadaveric specimens and assessed in a blinded manner using the ASSET score. (Arthroscopy Surgical Skill Evaluation Tool). Between the two tests, the control group performed no arthroscopy and received no training, the cadaveric group received four hours of practice on cadavers and the VR simulator group underwent four hours of training on the ArthroSim VR Simulator. The cadaveric-trained group showed the greatest improvement, with a mean difference in ASSET score of +4.27, +1.92 and -0.4 for the cadaveric, VR simulator and control groups respectively (p=0.04 for between group comparison). Residents in the cadaveric group showed a more rapid improvement compared to the VR simulator group, with a mean improvement of 1.1 ASSET points per hour of training vs 0.5 ASSET points in the VR group. The authors performed a cost estimation which showed that despite this the VR simulator may be more cost effective than cadaveric simulation if used at least 300 hours per year (this was institution specific so may not be widely generalisable)²⁵⁶.

Gottschalk et al²⁵⁵ measured the effect of simulation modality on the accuracy of cervical lateral mass screw placement by randomising 15 PGY1-5 residents to three

groups; cadaveric training, low-fidelity training (sawbones) or no training. The cadaveric and low-fidelity group performed drillings on the respective simulator modalities and received 3D navigational feedback on the intended drill trajectory. Baseline performance was measured on cadavers before randomisation, and they were re-tested following training. The final product was analysed by a single blinded expert. The results showed that both the cadaveric and low-fidelity simulator groups improved compared to the control group ($p < 0.0001$). The low fidelity group had a modestly higher (not statistically significant) post intervention improvement than did the cadaver group (aggregate mean distance from perfect screw angle of -8.2 and -7.2 degrees for low fidelity and cadaver groups respectively, where a negative difference is indicative of improvement). The authors conclude that training with 3D navigation 'significantly improved the ability of orthopaedic residents to properly drill simulated lateral mass screws' and that perhaps VR simulation offered a more cost effective means to achieve this training effect than using cadavers²⁵⁵.

There were 16 cohort studies that measured change in learner behaviour. Five of these studies used a comparative design; one study compared behaviour change in CST vs low-fidelity training participants²⁵⁷, three compared inexperienced vs experienced participant performance²⁵⁸⁻²⁶⁰, and one compared within-subject performance change after a CST intervention²⁶¹. Eleven of the 16 level 3 cohort studies used a non-comparative, descriptive study design^{262-270 272 274}.

When comparing CST participants with low-fidelity (VR) simulation trained groups performing laparoscopic sigmoid colectomy, Le Blanc²⁵⁷ reported that technical skills scores after training (using OSATS) were superior in the VR trained group. Global satisfaction was better with the cadaveric model, and the authors concluded that the human cadaver was more difficult to perform well on but was better appreciated by the learners.

The primary objective of the three studies that compared performance between inexperienced and experienced participants was to establish construct validity of the simulator and/or the assessment tools used in the studies.

Mednick et al²⁷⁵ compared the performance of 11 inexperienced ophthalmology residents (PGY 2) with 11 experienced ophthalmic surgeons during a single performance attempt at corneal rust ring removal. The results showed that the mean procedure time was longer in the inexperienced group (187 vs 117 seconds), that final product analysis scores were similar between the groups (as measured by the rust removal percentage) and there were no complications (perforation).

Mackenzie et al²⁵⁹ compared pre-training, immediate post-training (less than 4 weeks) and delayed interval performance (12-18 months post training) of lower extremity vascular exposure, repair and fasciotomy by 40 surgical residents (PGY3-6+) and 10 expert traumatologists. Their results showed overall longitudinal improvement with increased knowledge, correct procedural steps and decreased errors from 60 to 19% after training (which was independent of training level). Trauma Readiness Index scores for experts was significantly different compared to the trained residents before, immediately training and at delayed interval assessment. The authors concluded that interval experience rather than time since training affected skill retention up to 18 months later²⁵⁹.

Zirkle et al²⁵⁸ compared novice with experienced resident performance at a single attempt at cortical mastoidectomy on a cadaveric temporal bone. Performance was measured by two blinded independent experts using Global Rating Scale (GRS), Task-based Checklist (TBC), final product analysis (FPA) and expert opinion (EO). This study was atypical in that the primary objective was to validate these assessment tools for training in a cadaveric laboratory setting, rather than to test the impact of the cadaveric training on the learners' behaviour. The results showed that FPA does not correlate with trainee experience, but GRS and TBC scores did.

Wong et al²⁶¹ measured the within-subject performance change in saphenous venous cutdown performance, where the study participants acted as their own controls by assessing change in performance longitudinally. The participants were a cohort of 9 PGY2-4 residents undertaking a cadaveric lab dissection course. The procedure was attempted twice on days one and seven of the course, and assessed each time by two independent observers. Immediate feedback was given to the participants. There was improvement seen between the two attempts, shown by decreased mean incision size (4.6 vs 3.4cm, $p<0.05$), decreased time to completion of procedure (360 vs 232 seconds, $p<0.05$) and the decreased frequency of complications (37.5% to 0%, $p<0.05$).

Of the 11 non-comparative, descriptive cohort studies, all reported improvement in participant performance following a CST intervention, using a variety of assessment measures such as GRS, FPA and operational metrics.

Level 4: Results – studies that objectively measured educational impact by change in patient outcome

Only one study of 51 included in this review assessed the impact of the cadaveric training intervention on real-world patient outcomes. Martin et al²⁷³ recruited eight PGY1 surgical residents, and measured their baseline performance in three ‘core’ invasive skills (endotracheal tube insertion, chest tube insertion and venous cutdown) in the workplace setting during the first three months of their rotation. The residents then participated in two cadaveric laboratory sessions, held three weeks apart, using a competency based approach, where 100% procedural competency was the goal. Skills were repeated, with feedback, until they were performed successfully. Task failure was defined as an ‘inability to perform the task correctly or within 120 seconds’. Performance was again measured in the workplace, on real patient procedures, after the training. The results showed that all residents passed the skills immediately after instruction. At the second cadaveric session at three weeks, 8/8 participants passed VC and CTI, and 6/8 passed ETI. Changes in performance over time for both intervals (baseline-to-training, and training-to-three week post testing) were statistically significant ($p<0.05$). The complication rate for all skills

decreased significantly immediately and at three weeks after instruction ($p < 0.02$). Initial trauma resuscitation after training in 80 patients treated by these residents decreased from approximately 25 to 10 minutes²⁷³. The authors conclude that residents' skills rapidly improve with competency-based instruction (CBI), skills learnt through CBI in the laboratory can be translated to and sustained in the clinical setting, and that CBI yields competent residents who perform skills rapidly and with minimal complications. This study to date represents the highest level of evidence of patient benefit for cadaveric simulation training.

3.4.6 Cadaveric Models used for training

A summary of the prevalence of use of the different available types of cadaveric models in the included studies is shown below in table 5.

Table 5. Cadaveric models used in the review studies

Cadaveric Model	Number of Studies
Perfused/reconstituted 'live' model	3
Fresh	12
Fresh-Frozen	12
Thiel Cadavers	2
Formalin-Fixed	1
Not Stated	21
Total	51

Three studies^{184 239 272} used innovative techniques to perfuse or reconstitute cadaveric material for training purposes, to improve the fidelity of the simulation. These studies were all in the field of neurosurgery, and involved cannulation of the great vessels of the neck to allow pulsatile perfusion of cadaveric heads with an artificial blood substitute to simulate management of intracranial bleeding scenarios. All three studies reported very high learner satisfaction with the models, and recognition

of the opportunity that ‘live’ reconstitution offers for overcoming the oft-cited disadvantage of conventional, non-perfused cadaveric material in not simulating bleeding.

Fresh cadavers were used in 12 of the reviewed studies^{220 221 228 232 241 248 253 257 259 260 270 273}. These offer the most authentic tissue handling fidelity as they have not been subject to the tissue-altering processes of traditional chemical preservation²⁷⁶. Fresh cadavers do however have the significant disadvantage of rapid deterioration, and therefore there is a short time window for their potential use, which poses logistical restraints on their use in training. Use of fresh cadavers for simulation training relies on a regular, local system of body donation bequests, as they are typically used within 48 hours of the donor’s death, and no longer than 7 days.

Fresh-frozen cadavers were used in 12 of the reviewed studies^{222 226 231 235-237 243 247 250 251 261 262} and are a popular alternative to fresh material. The cadavers are non-exsanguinated, they are washed with antiseptic soap and frozen to -20 degrees centigrade within a week of procurement²⁷⁷. Usually about 3 days before use, they are gradually thawed at room temperature, thus retaining the realistic tissue handling characteristics on thawing that are so prized in cadaveric simulation. Fresh-frozen cadavers also have the considerable advantage of being able to be re-frozen and thawed again at a later date, thus permitting re-use across multiple training interventions and maximizing cost-efficiency²⁷⁷.

Soft-fix ‘Thiel’ embalming techniques were used in two studies^{238 263}. This technique for cadaveric preservation has been developed over the last 30 years, and was borne from a movement towards seeking a method of body preservation that preserved tissue handling, enabled longevity of specimen use, and avoided the known environmental and occupational health-risks associated with exposure to formaldehyde²⁷⁸. The organs and tissues of Thiel embalmed cadavers retain their flexibility, and the colour of the tissue remains similar to that seen in vivo, which is particularly important when simulating surgery.

One study²⁴⁶ used formalin-fixed cadavers. Formalin has the advantages of being inexpensive, easy to use and widely available, with a long history of successful use

in preserving cadavers for the purposes of teaching anatomy²⁷⁸. It does however lead to significant changes in the colour, strength and tissue handling characteristics of the cadaveric material¹⁷⁵, which may limit its usefulness in surgical training.

Sharma's study did not evaluate the fidelity of the simulation or discuss the rationale or impact on the educational value of the training as a result of using formalin fixated material, and as yet there has been no within-study comparisons of the utility of different types of cadaveric material for simulation training. Twenty one studies did not provide information about the type of cadaveric material used in the training intervention.

3.5 Discussion

3.5.1 Summary of findings

The aim of this chapter was to systematically describe and evaluate the current evidence base for the use of cadaveric simulation in postgraduate surgical training. I found 51 studies that met the inclusion criteria, involving a total of 2002 individual trainees across 69 cadaveric surgical training intervention exposures. The specialties with the highest research activity into cadaveric simulation are general surgery and T&O, which is not surprising as they are the two largest surgical specialties. The majority of the studies were conducted in the USA and UK.

As described above in the results section, a wide range of study designs were used. Eight of 51 studies were randomized controlled trials and six were parallel cohort studies. The majority of studies (37 of 51) were non-comparative descriptive studies (OCEBM level 3).

In the analysis I stratified the studies by Kirkpatrick level, to assign a hierarchy to the available evidence with respect to the measured educational impact of the cadaveric simulation intervention. Almost half of the included studies (43%) assessed the impact of the cadaveric training intervention using subjective measures only, which represents the lowest level of the hierarchy of impact assessment in educational research (Kirkpatrick level 1 – table 4a). The second most prevalent

category was studies measuring the impact of the training intervention by learner behaviour change (45%) (Kirkpatrick level 3 – table 4c). Of these, seven were RCTs and 16 were cohort studies.

Only one study²⁷³ achieved the highest level of Kirkpatrick's hierarchy (level 4 – table 4d) by measuring a change in patient outcome as a result of a cadaveric training intervention.

There is some reasonable evidence from three RCT's that cadaveric simulation is superior to no training²⁵⁰⁻²⁵², yet the arguably more important question of whether it is superior to low-fidelity equivalents is less clear. This is relevant because CST is expensive, and therefore needs to have clearly demonstrable pedagogic superiority over cheaper alternatives to justify the additional cost. The four RCT's²⁵³⁻²⁵⁶ that addressed this area in comparing CST with low-fidelity alternatives reveal a mixed picture. Two trials showed superiority of CST^{254 256}, one showed equivalence with low-fidelity bench model training²⁵³ and one showed cadaveric simulation training was inferior to the bench-model alternative²⁵⁵.

3.5.2 Methodological quality of the studies

The mean overall MERSQI had a clearly linear relationship with the Kirkpatrick level, the higher the level of evidence of impact, the better the study methodology (table 3).

The MERSQI scores ranged widely (3.5-14 out of a maximum score of 18). Previous work has shown that studies with a MERSQI score >10.7 were indicative of methodological rigor, and were significantly more likely to be accepted for publication²¹⁹. Of the studies in this review, 55% had a MERSQI score less than 10.7, and therefore it can reasonably be stated that the majority of the studies included in this review could be considered to have weak methods.

There were a variety of methodological problems amongst the included studies that were noted by both me and AWC in our independent appraisals. These included; a preponderance of single-site studies, small samples (with inherent risks of under-powering), lack of randomisation, lack of a comparator group, inadequate reporting of basic descriptive statistics, and a heavy reliance on a pre-test/post-test strategy, which is known to risk over-estimating the observed effect size of the training²⁷⁹.

It is inevitable, and understandable, that the known challenges of delivering cadaveric simulation including expense and dependence on specialist facilities may go some way to explain the observed problems with sample size and the lack of multi-site studies.

There were also several studies that reported complex mixed-modality training interventions, which makes assessment of the cadaveric component of the training in isolation impossible. In 21 of 51 studies there was also inadequate description of the cadaveric models used in the training intervention, which makes the generalisability of results pertaining to the face and content validities of the training impossible to assess.

3.5.3 Strengths and Limitations

A strength of the work presented in this chapter is that it is the first comprehensive, systematic appraisal of the evidence for cadaveric simulation in postgraduate surgical training. By structuring the evidence around an accepted framework for measuring educational impact (Kirkpatrick's hierarchy) and using widely validated methodology scoring tools (MERSQI and OCEBM), the review gives a clear and reproducible picture of the state of the evidence base and makes the deficiencies obvious.

A weakness is that the wide heterogeneity of the included studies precluded the application of pooled meta-analysis, and for this reason I chose to do a qualitative,

narrative synthesis. I did not access the ‘grey’ literature, and by excluding non-English language studies publication bias cannot be ruled out.

3.5.4 Areas for further work

Half of the studies were published in the last four years, which reflects the recent explosion in popularity of cadaveric simulation training. Despite its clear popularity and attraction as a surgical training tool as measured by subjective means, there remains a lack of high quality evidence that there is retention and translation of skills learnt in the cadaveric laboratory into the real-world operating theatre. A major, ongoing challenge within the educational research community is in demonstrating effective, sustained changes in learner behaviour and ultimately improved patient outcome following a training intervention²⁷⁹. These challenges are particularly acute and difficult to overcome in the field of cadaveric simulation research because of the cost and infrastructure demands in providing the training as outlined above, and are reflected in the results of this review.

3.6 Conclusion

The results of this chapter show that there is an abundance of relatively low-quality evidence showing that cadaveric simulation is a highly popular tool for training surgeons, and may induce short term behavioural change in the simulation training laboratory when measured by objective means. There is a lack of evidence of skill retention longitudinally, of transfer of skills following training into the live operating theatre, and of potential patient benefit. High quality, randomised studies using Kirkpatrick level 4 outcomes are required to address this deficit.

3.7 Reflections

In this chapter's reflections, I will consider what I have learnt in the process of undertaking a systematic review.

Prior to the start of this work, I had read and cited systematic reviews but had never undertaken one myself, and did not have a clear understanding of the process involved. With the guidance of my supervisors, I read up on the methods for conducting a systematic review, with a particular emphasis on the various methods for conducting narrative reviews in medical education, and for appraising methodological rigor in educational research. I learned about the PRISMA guidelines and the importance of method transparency and standardized reporting in systematic reviews.

I initially under-estimated the level of organisation required to undertake a systematic review. I quickly developed some good habits in this respect. I learnt how to generate a electronic resource database profile enabling me to save the search strategy. I began to use professional reference management software at the start of this chapter which helped keep everything organised, and quickly became indispensable once the scale of the work became apparent. I entered the study data into a spreadsheet during abstraction, and was careful to manage version control when it was being viewed and edited by more than one person.

I discovered in the final stages of undertaking this work that software packages exist for managing reviewer input in a systematic review, and which allow reviewers to work remotely to screen and assess studies. I will certainly be using this for the next review.

I was really pleased that the finished review has been published in the British Journal of Surgery, and that we subsequently received some correspondence about the article which I responded to in print.

Chapter 4: The current status of simulation provision in UK and Republic of Ireland Trauma & Orthopaedic specialist training programmes: a national survey

In this chapter, I will investigate the current provision of simulation in T&O training programmes, to understand the landscape of what facilities are available, the sources of funding for simulation training, and the barriers to access/delivery. I will also explore current research activity into the impact of simulation delivered within training programmes.

Declarations:

This chapter has been submitted for publication;

James HK, Gregory R, Tennent D, Pattison GTR, Fisher JD, Griffin DR. Current provision of simulation in the United Kingdom and Republic of Ireland Trauma & Orthopaedic specialist training: A National Survey. Submitted to Bone & Joint Open.

This chapter has been presented;

James HK. Current provision of simulation in United Kingdom and Republic of Ireland T&O Specialist Training. Keynote lecture in education session, British Orthopaedic Association Annual Congress, Liverpool, September 2019

The survey was administered by Mr Rob Gregory, SAC Chair and owner of the Performance and Opportunity Dashboard

Map design was done by Mrs Yessica Diez-Davies from my hand drawings

4.1 Introduction

There appears to be increasing interest in, and provision of, simulation training opportunities for T&O. Simulation courses are regularly advertised to trainees via the British Orthopaedic Trainees Association (BOTA)²⁸⁰ and the Surgical Royal Colleges²⁸¹. Currently, these advertised courses are at cost to trainees and are optional adjuncts to training.

With cadaveric simulation training specifically, most courses are funded wholly²⁸² or partly²³¹ by industry, and are often aimed at consultant surgeons, to learn to use a new implant or piece of equipment. These courses therefore have a marketing agenda beyond simply learning surgical skill. Whilst it undoubtedly makes financial sense for a surgical training centre/cadaveric laboratory to have a portfolio of commercial courses, it is unclear what the current provision is for surgeons-in-training.

In addition to cadaveric simulation, there is some evidence of current research activity into other simulation modalities within T&O, including the development²⁸³ and application²⁸⁴ of virtual reality, computer-based simulation. It is not clear whether this activity is confined to the research setting, or is part of a wider provision within training.

There is currently no comprehensive overview of the provision of simulation within T&O training programmes, or of research activity in simulation for T&O.

The aims of this chapter are to;

1. Map the current provision of simulation training within UK and Republic of Ireland (RoI) T&O specialist training programmes
2. Characterise the types of simulation offered to trainees by stage of training
3. Identify the sources of funding for simulation in T&O training programmes

4. Identify the barriers to providing simulation training
5. Describe how the educational impact of simulation training is being measured within training programmes.

4.2 Methods

4.2.1 Survey Development

My primary aim in doing the survey was to map the current provision of simulation training within UK and RoI T&O specialist training programmes.

I developed questions to describe, by programme and region; the resources available for simulation training, sources of funding, timetabling/logistics of provision within training programmes, the types of simulation offered at each stage of training, and current research activity measuring the educational impact of simulation training.

The survey questions are shown in appendix 11.2.

4.2.2 Participants

Training Programme Directors (TPDs) for each of the 30 individual T&O training programmes in the UK and Republic of Ireland (RoI).

4.2.3 Survey Administration

A web-survey of all TPDs of UK and RoI T&O specialist training programmes is undertaken annually by the Specialist Advisory Committee (SAC). This is administered via a commercially available platform (Survey Monkey Inc., San Mateo, California, USA). The results of this annual survey, which is mandatory,

provides quality information on training programmes for the T&O Performance and Opportunity dashboard.

My survey questions were embedded within the ‘simulation’ domain of the 2019 annual TPD web-survey. I thought that by embedding my questions within the annual TPD survey, rather than designing and implementing a separate survey, a high response rate could be achieved.

I designed the survey questions for clarity, brevity and appropriateness to generate data to fulfil the research aims. Stemmed, multiple choice questions were used to elicit the required detail and be usable within the web-survey.

The SAC had previously stated²⁸⁵ there was a need for a national audit of simulation provision in T&O. I was invited to undertake this work, and the timing and scope aligned nicely with the development of this thesis. A fruitful collaboration with the Dashboard owner/SAC Chair (RG) and BOA Simulation Lead (Professor Duncan Tennent) resulted. The results of my input to the annual survey became the results of the BOA audit and are the basis of this chapter.

Ethics approval was already in place for the dashboard and additional permission for my input was not required. Whilst participation in the annual survey was mandatory for the respondents, my questions were not and were signposted as being part of a research project.

4.2.4 The T&O Performance and Opportunity dashboard

Until 2018, postgraduate training in T&O was assessed using three outcomes; GMC survey, JCST survey and SAC liaison reports. Most of this data is trainee-centric making global assessment of a training programme difficult, and lacks detail on specific training programme activity and outcomes. The T&O performance and opportunity dashboard was developed by the specialist advisory committee to help achieve parity between training regions, and to identify areas for improvement in

training²⁸⁶. Data for the dashboard is collected annually by web survey as described above, and enables the auditing of quality in UK and RoI T&O training programmes.

The survey is across 9 domains;

1. Fellowship of the Royal College of Surgeons (FRCS) exam pass rates
2. GMC Survey results
3. Research opportunities and output
4. Simulation
5. Enhanced induction
6. Subspecialty access
7. Professional support
8. Education
9. Study leave funding

The results of the dashboard survey are used to identify areas of excellence and weakness in training programmes, and are made available to prospective trainees to help them inform their training programme choices at ST3 interview²⁸⁶.

4.2.5 Data Analysis

The data from my input to the simulation domain was retrieved by RG from the dashboard and sent to me electronically. I undertook a descriptive analysis, where responses are presented as counts and percentages, and also presented visually as geothermal heat maps to show regional trends.

4.3 Results

4.3.1 Demographics

Twenty-eight of the 30 eligible TPDs completed the survey, giving a response rate of 93%. The territory boundaries of the 30 UK and RoI training programmes are shown in figure 12.

There was no map available that showed the geographical distribution of the training programmes. Maps were therefore created by plotting the hospitals within a given training rotation on Google maps (Google, Alphabet Inc., Mountain View, California, USA) and hand drawing the boundaries around them. The map was then sent to a graphic designer (YDD) who recreated them electronically.

The geographical size of the programmes varies widely, and is independent of the number of trainees hosted within each programme. There are seven programmes in London, and there is considerable overlap between the rotations amongst the London teaching hospitals. I took a pragmatic, best-fit approach to determine the training programme boundaries in London, to allow for clear visual representation of the data.

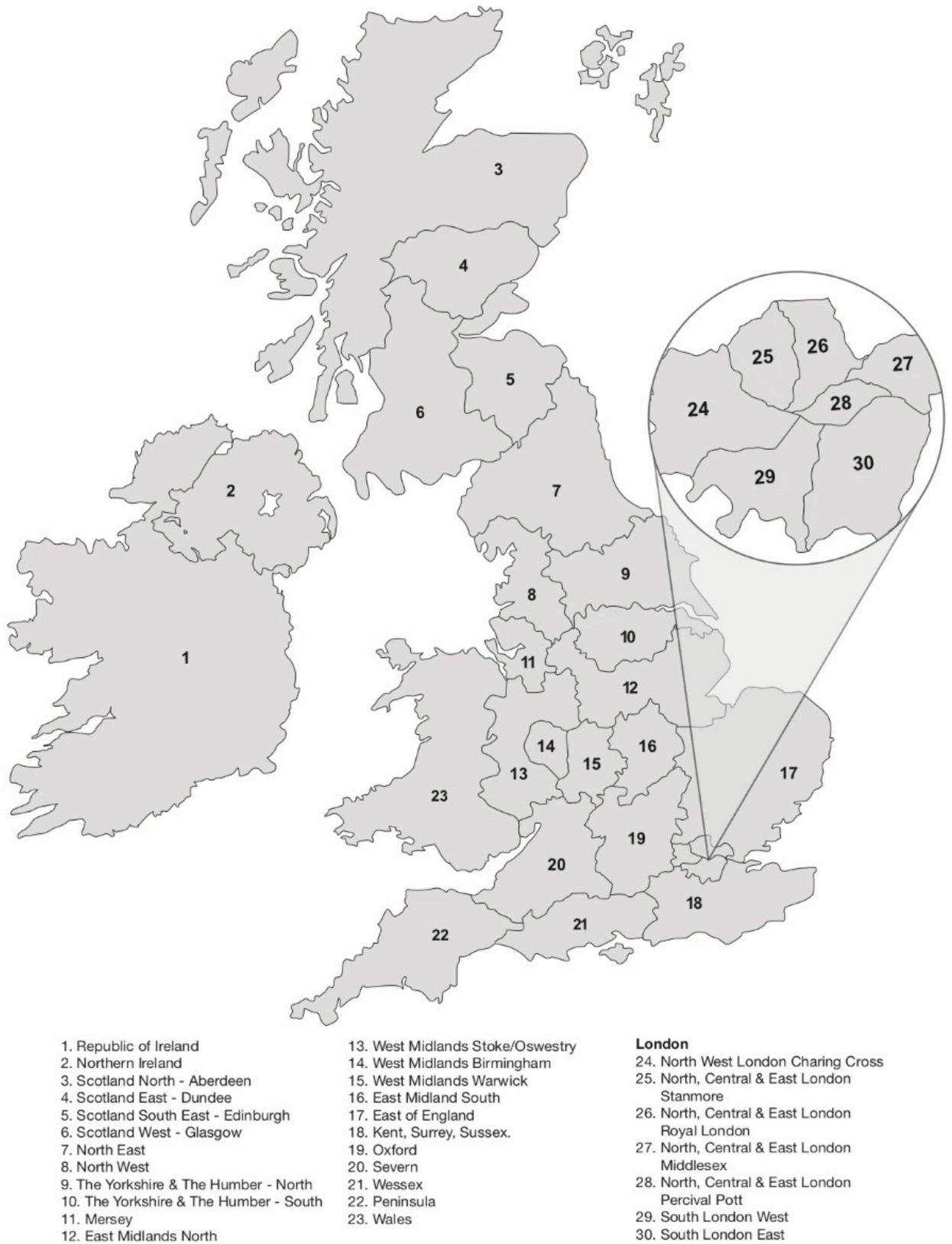


Figure 12. Geographical boundaries of the UK and ROI T&O training programmes

4.3.2 Resources for simulation

Twenty-seven (96%) of programmes reported facilities for clinical case management, e.g. clinical skills teaching rooms within hospitals. Just over half (54%) had access to non-technical skills simulation, delivering modules related to situational awareness, decision making, communication and teamwork. In terms of technical surgical skills teaching, 82% had access to cadaver-based surgical procedure simulation and 64% of programmes had access to arthroscopy simulation. Less than one-third (29%) of programmes had access to a simulated operating theatre environment (figure 13).

Four programmes (14% of the total, three in London) offered access to simulation facilities to trainees out-of-hours on an 'as required' basis.

I classified simulation resources by programme as very low, low, moderate, high and very high, and the geographical distribution of simulation facilities are shown in figure 13 with classification descriptors. There was no clear regional pattern to provision of facilities for simulation.

Training (on-line, on-paper, or face-to-face) was available for the simulation trainers in five regions (shown in figure 13 by the dotted shading). I cannot see a relationship between this provision and the overall quality of the simulation facilities within the training programmes. The opportunity to be involved in simulation training was made available to all trainers in just under half (46%) of programmes.

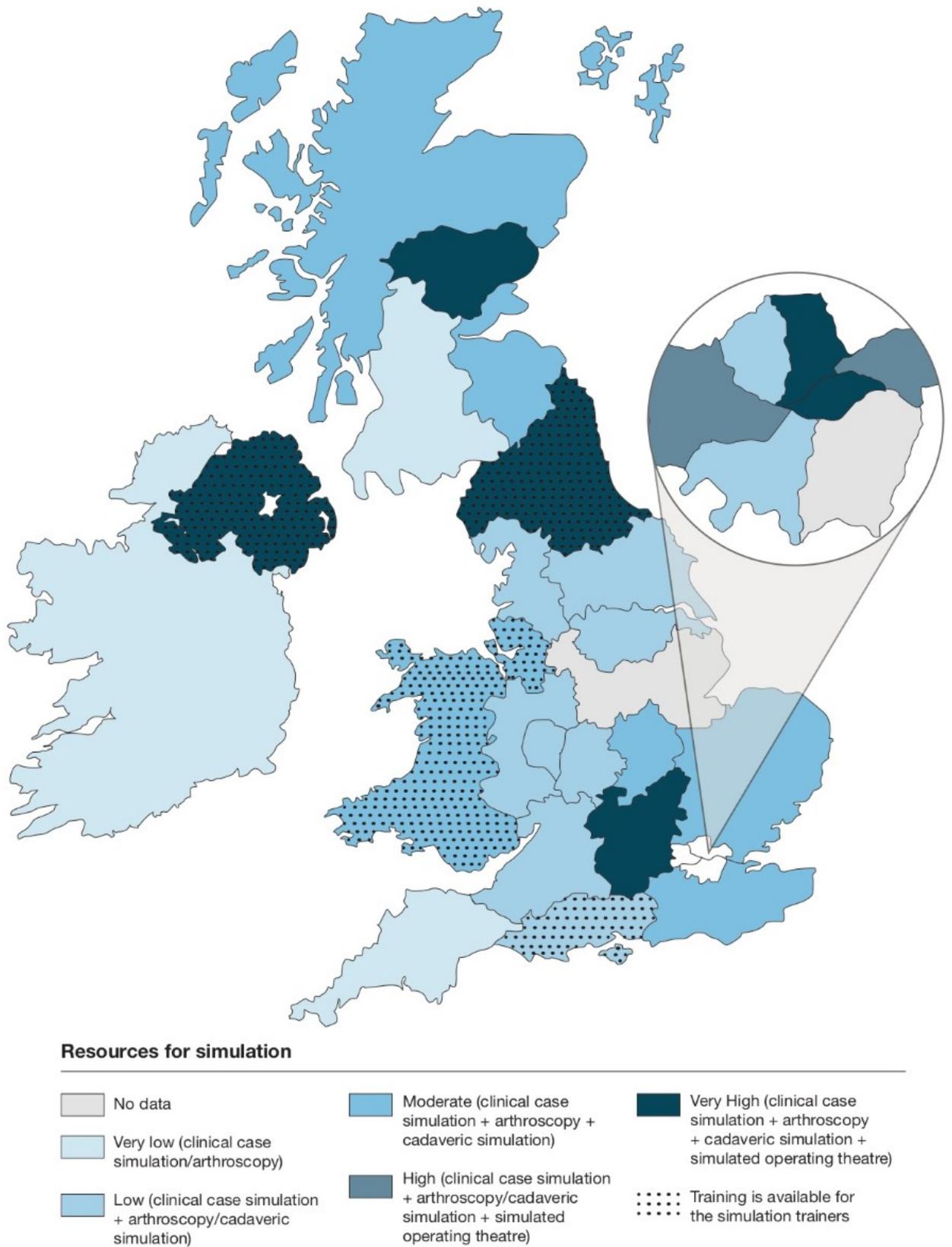


Figure 13. Map showing resources for simulation by programme

4.3.3 Funding sources

Less than half (43%) of programmes reported receiving centralised funding for simulation provision from Health Education England and/or the Postgraduate Deanery (figure 14). A third (32%) of programmes funded simulation provision locally, using either NHS Trust/Departmental budget (two programmes), top-slicing trainee study budgets or trainee self-funding (six programmes) or via a charitable foundation (one programme). Of note, two-thirds (64%) of programmes received industry or other commercial sponsorship for the provision of simulation.

No clear relationship was found between lack of centralised funding and reliance on industry sponsorship, and I did not seek quantitative estimates of funding in the survey. Programmes with centralised funding support for simulation were mainly clustered in the south of England (figure 14), and therefore there seems to be regional inequality in centralised funding for simulation provision in training.

Enquiries to HEE to ask why this was the case revealed the funding landscape for health education is incredibly complex with no straightforward explanation for this apparent phenomenon. A detailed understanding of this funding structure is beyond the scope of the thesis, and so I did not pursue my enquiry further.

Six programmes do not have funding information, as this was withheld by the survey respondents.

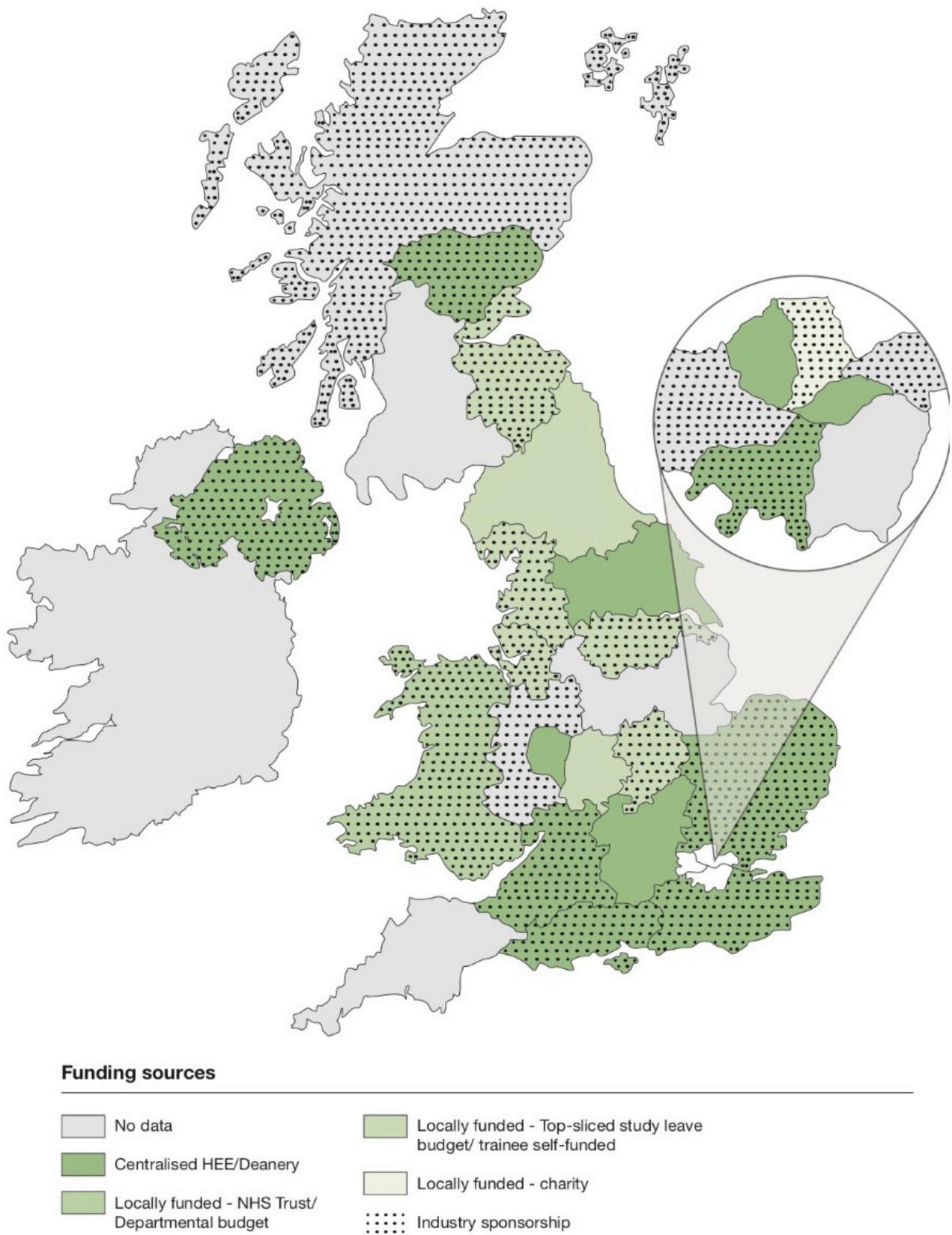


Figure 14. Map of funding sources for in-programme simulation provision

4.3.4 Barriers to provision

I wanted to understand the barriers to delivering simulation in training programmes, and to see if this explains regional variation. In the survey I asked the TPDs about the obstacles they perceived to be facing in delivering simulation training (figure 15). Forty percent reported barriers; six programmes reported a lack of facilities, two had funding issues, two reported logistical difficulties related to timetabling, one reported lack of available faculty, and one reported lack of trainee enthusiasm for simulation training (figure 15).

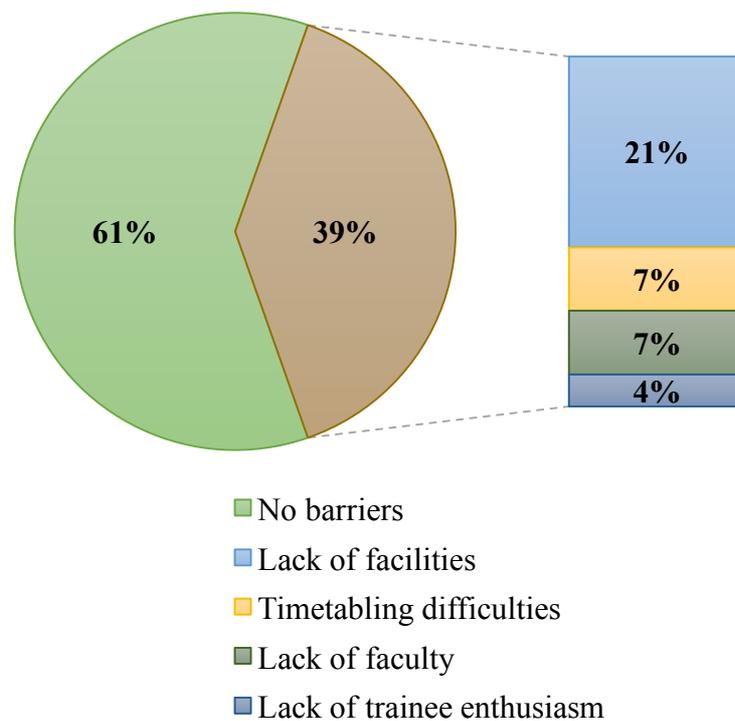


Figure 15. Barriers to delivering simulation training within programmes

4.3.5 Formal provision of simulation in the timetable

I asked the TPDs whether they formally timetabled simulation into the teaching programme. I felt this was an important question, as it is likely to impact the accessibility by trainees, as it avoids the problem of scheduling conflicts.

Simulation was formally timetabled in 19 (68%) of training programmes, and was used as part of an enhanced induction programme in nine (32%). There was a clear relationship between the provision of a simulation enhanced induction programme and formal timetabling of simulation, with only one programme offering the former without the latter. When represented visually on the map (figure 16), simulation was more likely to be formally provided in the timetable in England and Wales than in Scotland, Northern Ireland or the Republic of Ireland.

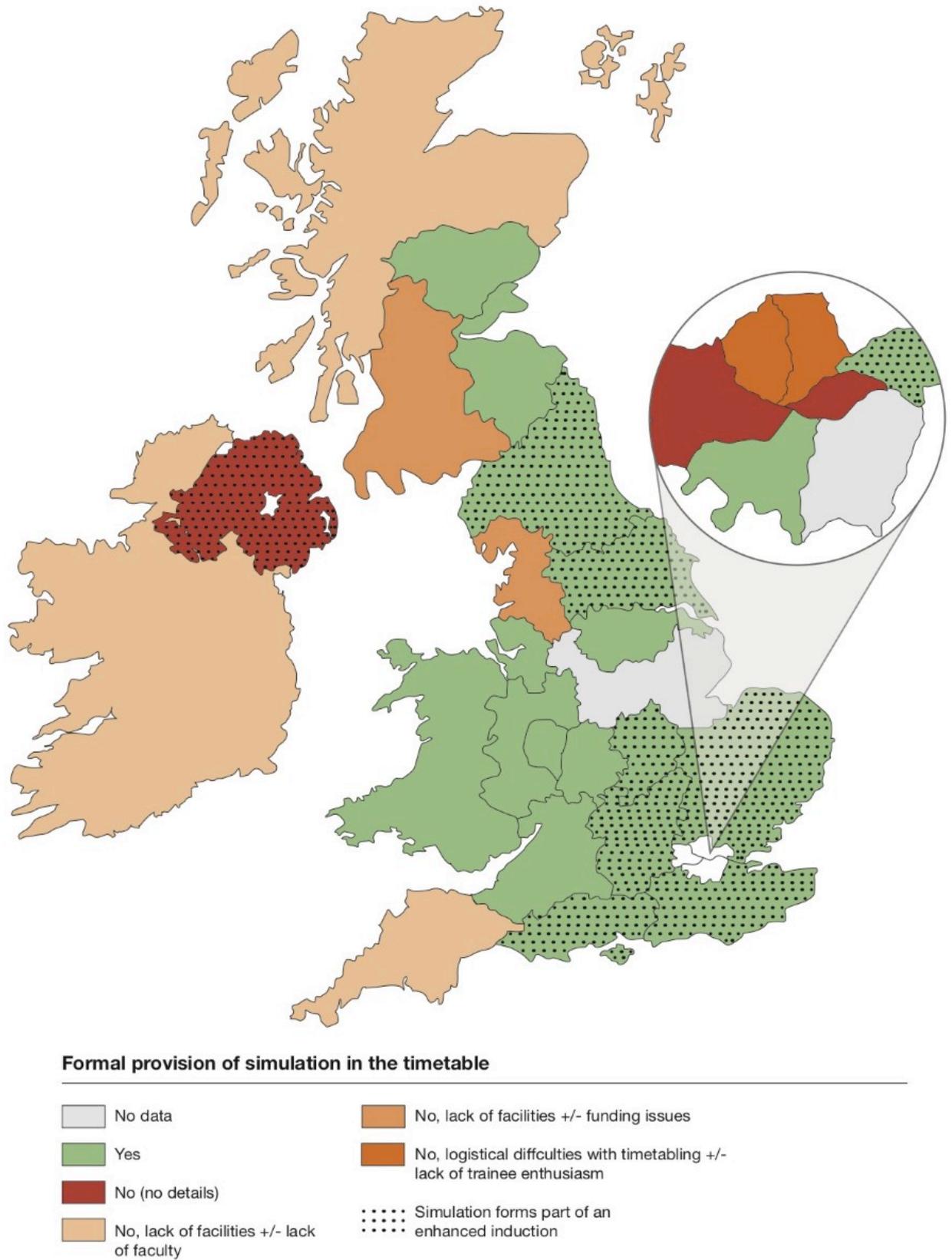


Figure 16. Map of formal provision of simulation within the teaching timetable

4.3.6 Types of simulation offered by stage of training

I was interested to know if the provision of simulation varied by stage of training, as the learning needs of trainees differ by stage. I considered the 'mid-stage' of training as ST3-5, and the 'late-stage' as ST6-8.

For the purposes of the survey, I defined low fidelity simulation as box trainers or Sawbones (Sawbones Europe AB, Malmoe, Sweden) or equivalent. Moderate fidelity was defined as virtual reality (VR) or animal tissue. High fidelity was defined as human cadaveric simulation involving surgical approaches/dissection only. Ultra-high fidelity simulation was defined as human cadaveric simulation with the addition of implants, instruments and the ability to perform the whole procedure on the cadaver.

The maps showing the types of simulation offered at the middle and late stages of training are shown in figures 17 and 18. As would be expected, the degree of fidelity of simulation offered by stage broadly correlates with the facilities available for simulation within the respective programmes.

There appears to be a trend towards less simulation provision at the later stages of training, with two programmes offering no simulation provision at ST6-8, and three programmes offering high and ultra-high fidelity simulation to ST3-5 only.

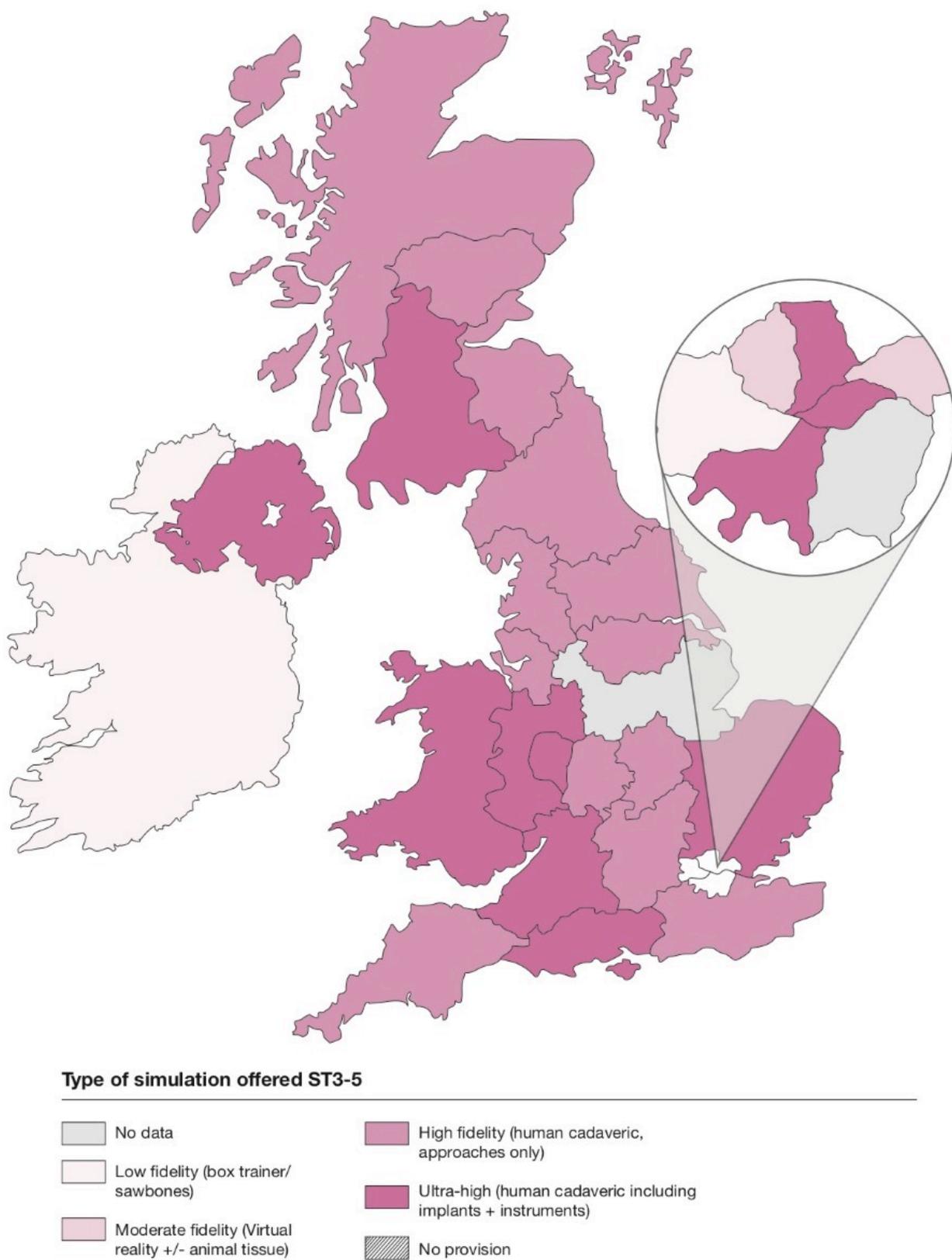


Figure 17. Map of simulation provision at the mid-stage of specialist training

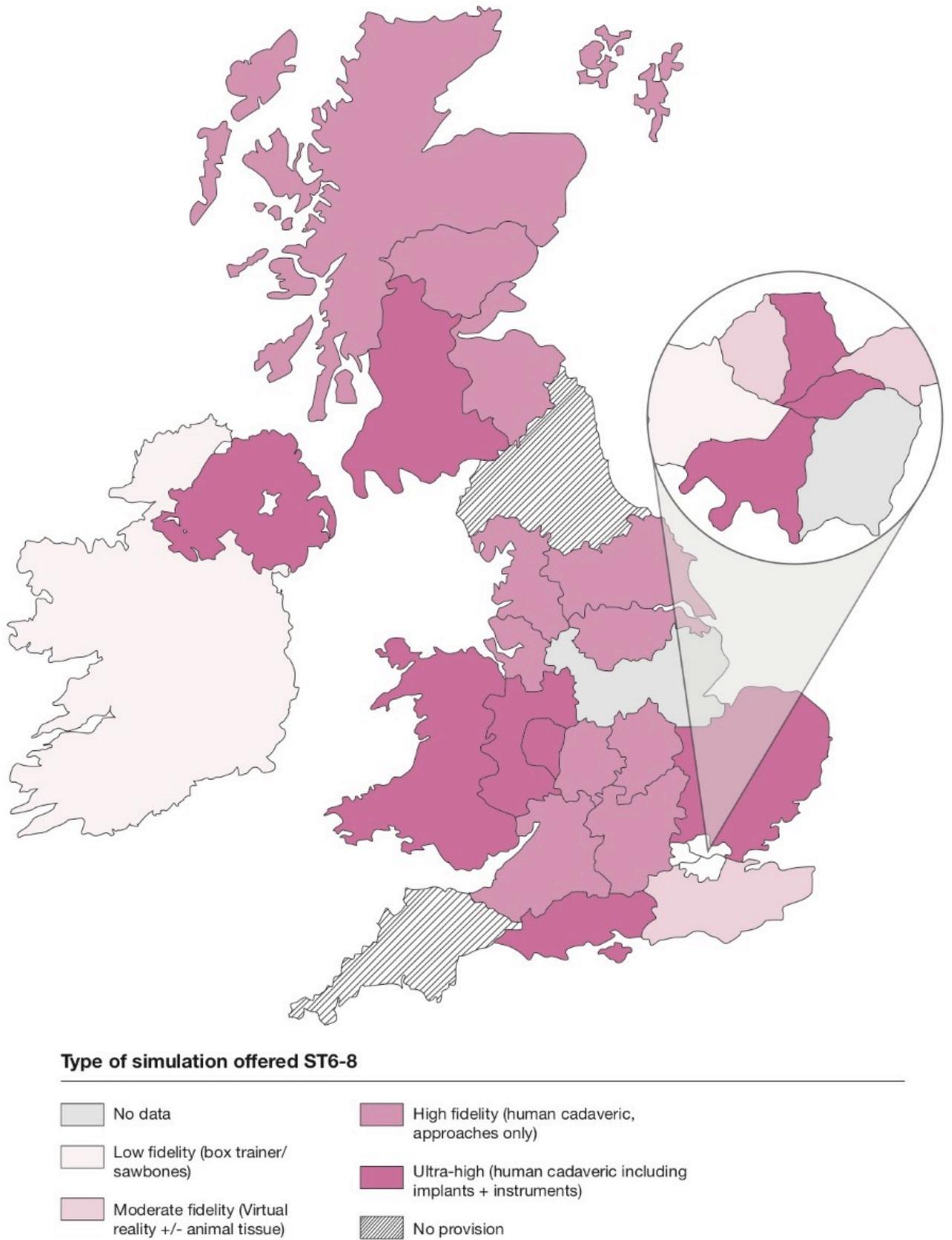
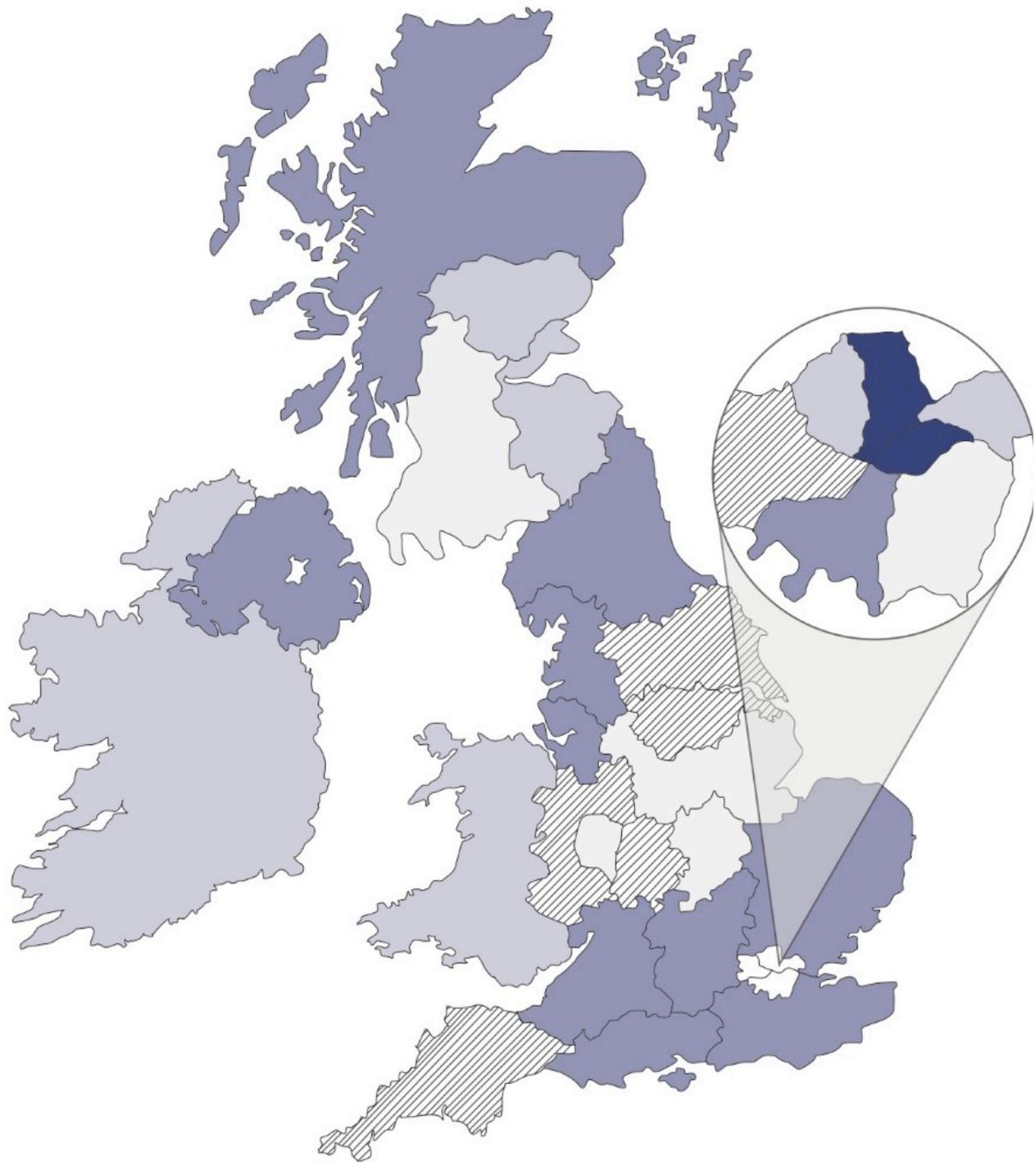


Figure 18. Map of simulation provision at the late stage of specialist training

4.3.7 Measuring the effect of simulation training

I stratified activity to measure the educational impact of simulation training according to Kirkpatrick's hierarchy (chapter 2, figure 8), as this is a widely accepted framework for classifying the educational outcomes of a training intervention. Six programmes (21%) did not measure the impact of their in-programme simulation provision. Six programmes (21%) measured the effect of simulation training on trainee opinion, using pre-/post-training questionnaires to assess subjective metrics such as procedural confidence (Kirkpatrick level 1).

Eleven programmes (40%) assessed simulation provision using changes in trainee knowledge (Kirkpatrick level 2), through either bespoke post-training knowledge testing or the annual UKITE exercise²⁸⁷. Neither of these methods assess technical or clinical skill improvement from simulation nor skill transfer to the workplace. Two programmes in London report measuring the impact of simulation using subjective behavioural measures (Kirkpatrick level 3b). I thought it was striking that there were no programmes that assess the impact of in-programme simulation provision using objective behavioural measures (Kirkpatrick level 3a) or patient outcomes (Kirkpatrick level 4).



Measuring the effect of simulation training



Figure 19. Map of activity measuring the impact of in-programme simulation provision

4.4 Discussion

4.4.1 Provision is patchy

Orthopaedic educators face the considerable challenge of continuing to train high quality surgeons amid increasing clinical, financial, regulatory and time pressures on training²⁸⁸. As discussed in chapter 1, the traditional master-apprentice model of surgical training, with its reliance on a high volume of case exposure and a sustained mentor-mentee relationship¹⁰ has been rendered obsolete in the modern surgical healthcare environment⁷. This is due largely to increased service demands at the expense of training⁹⁴, shift based working patterns and short training rotations.

There is a growing evidence base supporting the use of simulation as an adjunct to training, which I have explored in detail in chapter 2. This evidence shows that the learning curve can be advanced away from patients⁸⁴, with inherent safety advantages²⁸⁹, and that skills learnt in the simulated environment can transfer to the operating theatre¹⁷⁴. There have been many studies demonstrating the face validity²⁹⁰, construct validity^{291 292}, feasibility²⁹³ and educational impact^{294 295} of various orthopaedic simulators for both open and arthroscopic surgery, ranging from low fidelity, low cost box-trainer type models through to ultra-high fidelity cadaveric simulation.

The survey results show that overall simulation provision is highly variable across the 30 T&O training programmes of the UK and RoI. The availability of resources varies widely (figure 13), and I believe this is likely to be at least partly influenced by the geographic and financial relationship of training programmes with medical schools, where cadaveric wet-lab facilities are generally situated.

The funding landscape for simulation provision is complex and not explored in detail here. There was a tendency for simulation provision in the southern half of England to be described as centrally funded, the reason for which is unclear, and a widespread reliance on industry sponsorship (64%) was seen, seemingly independent of centralised funding status. Of note, four programmes reported that trainees were

required to self-fund simulation training. This raises obvious ethical concerns around equity of access to training opportunity, as some trainees may be disadvantaged through being unable to pay to access simulation training. Six studies reported funding simulation training by top-slicing trainees' study budget, which although fairer than asking trainees to self-fund, erodes their already limited study budget and may jeopardise access to other valuable training opportunities such as attendance at courses and conferences.

Barriers to formally integrating simulation in the timetable were explored, in order to understand the reasons behind regional discrepancies in provision of simulation. Barriers cited by TPDs in the survey included; lack of facilities (n=6), lack of faculty (n=1), funding issues (n=2), logistical issues with timetabling (n=2) and, surprisingly, lack of trainee enthusiasm (n=1). The majority of programmes (68%) had formally timetabled simulation provision, which is important for logistical considerations when planning training, and also assures the trainees of protected 'bleep-free' teaching time. Formal timetabling also removes the need for a separate study leave application process or annual leave usage to attend simulation training, and attendance rates can be monitored.

Type of simulation offered by stage of training and region is shown in figures 17 and 18. A disparity is seen between available facilities and reported provision in Glasgow, where ultra-high fidelity training is provided at both the mid- and late-stages of training but the resources for simulation is given as very low with no cadaveric facilities. One explanation is that cadaveric training is outsourced to another region, which we did not address in the survey. Less simulation provision was generally available during late stage training (ST6-8). This may be explained by the potentially greatest gains being obtained from simulation at an earlier stage of training, where the learning curve is steeper.

4.4.2 Strengths and limitations

A strength of this study is that it is the first attempt to map simulation in T&O training, and describe its current provision status on a national scale. A high response rate was obtained (93%), yielding a near complete overview of provision.

A limitation of this study is that I have focussed on establishing facts about provision at the expense of a more nuanced understanding of TPD opinion on the role of simulation in their training programmes. I have also not included trainees in this study who, as the direct beneficiaries of simulation training, should be central to discussions around provision. Despite the high response rate there are two training programmes for which we have no data. Provision of simulation is only one of several quality indicators relevant to assessing a training programme, and I can make no inferences as to the quality of individual training programmes based on these results.

4.4.3 There is a need for an evidence base

Despite the obvious appeal of simulation as a part-solution to the challenges of the modern surgical training environment, there is no evidence to date that simulation training in T&O benefits patients (Kirkpatrick level 4 evidence). T&O lags behind general surgery in its efforts to measure the educational impact of simulation training. In a systematic review¹⁹³ of skill transfer to the operating theatre after simulation training only one²⁹⁶ of 34 studies was from T&O. Research efforts to measure so called ‘transfer validity’ have been complicated by a lack of appropriately validated objective outcome measures²⁹⁷, and are largely restricted to the research setting. Only two studies^{296 298} have shown evidence of transfer validity following simulation training in T&O, both of these involve diagnostic knee arthroscopy. Arthroscopic procedures lend themselves more easily to objective measurement of skill transfer, as motion analysis can be used to objectively measure performance in the simulated environment, and subsequently the operating theatre. The measurement of transfer validity of open procedures is considerably harder²⁹⁷.

Simulation delivery can be costly, cadaveric training especially so¹⁷⁶. Until such a time that there is a high-quality evidence base showing that simulation training in T&O improves technical and non-technical skills that translate into the workplace for the benefit of patients, it is unlikely that simulation is going to be mandated for training or comprehensively funded by HEE.

The survey results show that despite some challenges there is currently wide-ranging simulation provision in training, much of it using sophisticated techniques such as cadaveric simulation. The use of innovative funding streams, and widespread efforts, albeit imperfect, to measure the educational benefit of simulation suggests a high level of enthusiasm for the delivery of simulation in T&O. These results reveal a promising foundation for the future of simulation delivery in T&O training, which I anticipate will continue to grow further with the implementation of the new curriculum in 2020. It is important that research activity continues to keep pace with the anticipated expansion of simulation provision, and can inform future developments in an evidence based manner.

4.5 Conclusion

There is currently widespread, but variable, provision of simulation in T&O training in the UK and RoI, which may expand further with the new curriculum. It is important that research activity into the impact of simulation training continues, to develop an evidence base to support investment in facilities and provision.

4.6 Reflections

After writing the introductory chapters of this thesis, and undertaking the systematic review of cadaveric simulation for postgraduate surgical training, I could see that there was a clear need for a detailed overview of current simulation provision within T&O training in the UK and RoI. This was important to understand what was currently going on nationally, and to provide context to the trial, situating it within current practice.

By a serendipitous set of circumstances, I came to be invited to contribute my research questions to the next round of the dashboard data collection cycle, the timing of which was just right for the project trajectory at that point. This was a particularly great opportunity for two reasons; firstly, it allowed me to ‘piggy-back’ onto an existing, established data collection process that would target my population of interest - training programme directors. I knew that dashboard submission was mandatory and so although my input was flagged as being for research and optional, I was hopeful that most of the TPD’s would complete my questions as part of their dashboard submission – and 93% of them did. This is probably a much higher response rate than I would have got if I had designed and administered my own survey independently of the dashboard.

Secondly, it allowed me to collaborate with, and get to know, some members of the Specialist Advisory Committee who design and steer training in the UK. This expanded my professional and research network, and I was subsequently invited to present the results of the survey as a keynote lecture at the British Orthopaedic Association Annual Congress in September 2019. This gave my work, and by extension me, exposure on a national stage, and generated some useful professional relationships.

I also valued the opportunity to work in a collaborative manner with the people who are the driving force behind UK and RoI T&O training reform, and it was satisfying on a personal level to know my work was at the forefront of curriculum redesign and would help inform decisions around simulation provision.

Chapter 5: Assessment of skill acquisition and operative competency in Trauma & Orthopaedic training

This chapter will be a systematic review of the current evidence for assessment of skill acquisition and operative competency in T&O training. In designing a randomised trial to measure the impact of cadaveric simulation training, careful consideration needs to be given to the appropriate primary outcome measure that is to be used.

Declarations

This chapter has been published;

James HK, Chapman AW, Pattison GTR, Griffin DR. Assessment of technical skill acquisition and operative competence in Trauma and Orthopaedic surgical training: A systematic review. JBJS Reviews. 2020, In Press.

Mrs Anna W Chapman was second reviewer and provided independent assessment of references for inclusion.

5.1 Introduction

As part of a widespread movement towards training quality improvement in T&O, there is an educational paradigm shift towards competency-based measures of performance²⁹⁹. There is a growing need to evaluate surgical skills objectively and systematically, and hence a drive towards developing more reliable and valid measures of surgical competence²⁹⁹⁻³⁰¹.

Several competency and skill assessment tools are currently in use in T&O training³⁰² and numerous studies have been performed to assess the ability of these various tools to assess surgical performance. A systematic appraisal of the evidence for use of these tools in training is necessary to make judgements about the appropriateness of the available tools as candidate outcome measures for my trial in this thesis. A systematic review will also be of value to the surgical education community and form part of the evidence base in the literature.

5.1.1 Appraising educational assessment tools

The first challenge is to decide on a framework for appraising the various assessment tools that are in current use in T&O. One of the most widely accepted methods of appraising assessment tools is Van der Vleuten's utility index³⁰³. The utility index appraises assessment tools across five domains which are described below in table 6.

Table 6: Van der Vleuten's Utility Index. Adapted from Bartlett et al³⁰⁴

1. Validity	The extent to which the skills claimed to be being assessed are assessed by the instrument
<i>Content Validity</i>	Describes the appropriateness of the variables measured by the assessment instrument ³⁰⁴
<i>Construct Validity</i>	Describes the effectiveness of the assessment instrument at differentiating between different skill levels ³⁰⁴
<i>Concurrent Validity</i>	Describes the extent to which the assessment instrument agrees with existing performance measures ³⁰⁴
2. Reliability	Describes the reproducibility of the results
3. Feasibility/Acceptability	The extent to which the instrument is usable by the target audience
4. Educational Impact	Consideration of the extent to which the instrument itself influences learning
5. Cost-effectiveness*	The extent to which the assessment instrument delivers value for money

*Cost-effectiveness has been excluded from this review as I felt that a detailed economic analysis of assessment tools was not relevant to the research question.

5.2 Aim

The primary aim of this review is to systematically evaluate the T&O surgical competency literature, and to report on the various metrics used for skill assessment in orthopaedic training, and their utility according to Van der Vleuten's index;

validity, reliability and impact on learning, and the evidence for strengths and weaknesses of the various tools.

5.3 Methods

I carried out this review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines³⁰⁵ and I registered it with PROSPERO (International prospective register of systematic reviews, CRD42019134657)²¹⁵.

5.3.1 Data Sources

I undertook a comprehensive literature search of MEDLINE, Embase and Google Scholar electronic databases. My search strategy was developed by collecting key words obtained from an initial scoping search. I kept the search deliberately broad to capture all studies in which a T&O surgical skill was assessed.

An example of the search strategy is reported below in table 7. Categories 1, 2 and 3 were combined using Boolean operators ‘AND/OR’ and results limited to Humans. I did not apply date or language limits. The last search was performed in May 2019. I removed duplicates, and retrieved the titles to screen for initial eligibility.

Table 7. Example search strategy

1. Competence.mp. OR assessment\$.mp. OR skills\$.mp. OR training.mp. OR performance.mp.
2. technical.mp. OR operative.mp. OR simulation.mp.
3. orthop\$.mp.
4. Combine 1 AND 2 AND 3

Syntax was adapted for different databases, as required.

5.3.2 Study Selection

Inclusion criteria:

1. Primary empirical research involving assessment of open or arthroscopic orthopaedic surgical skills
2. Assessment in simulated or live operating theatre environments.

Exclusion criteria

1. Non-empirical studies
2. Studies solely focused on patient or procedural outcome
3. Descriptive studies of a training intervention.

5.3.3 Title and abstract review

The search identified 2305 unique citations. I undertook initial abstract screening, where clearly irrelevant studies were excluded. 187 abstracts subsequently underwent independent screening by me and the second reviewer (AWC). 81 abstracts were rejected at screening because they were not empirical research, the study participants were either undergraduate students or fully trained surgeons, or non-technical skills were being assessed in the study. Studies that solely reported simulator protocol development or validation were also excluded at abstract screening. 106 studies were retrieved in full-text.

The reference lists of full-text articles were examined for relevant studies, and those found by hand searching were subject to the same eligibility screening process described above.

5.3.4 Data extraction and analysis

Data relevant to the review objectives was extracted into a structured form to ensure consistency. The data extraction form is attached in appendix 11.3. I did all the data extraction. The extracted data included; study aim, setting, assessment format, number and training stage of participants, skills assessed, assessment tool and/or metrics, assessment tool category, study results and ‘take home’ message related to the assessment tool. Assessment tools were classified by the type of method, and the following categories were defined; traditional written assessments, objective assessment of technical skill, procedure specific rating scale, individual procedural metrics, movement analysis, psychomotor testing and subjective assessments.

5.4 Results

5.4.1 Search results

The flow of studies through the review is shown below in the PRISMA diagram (figure 20). I assessed 106 papers in detail, and one was excluded at this stage as the participants were not surgical trainees. This was not clear from the abstract. 105 papers were therefore included in the review.

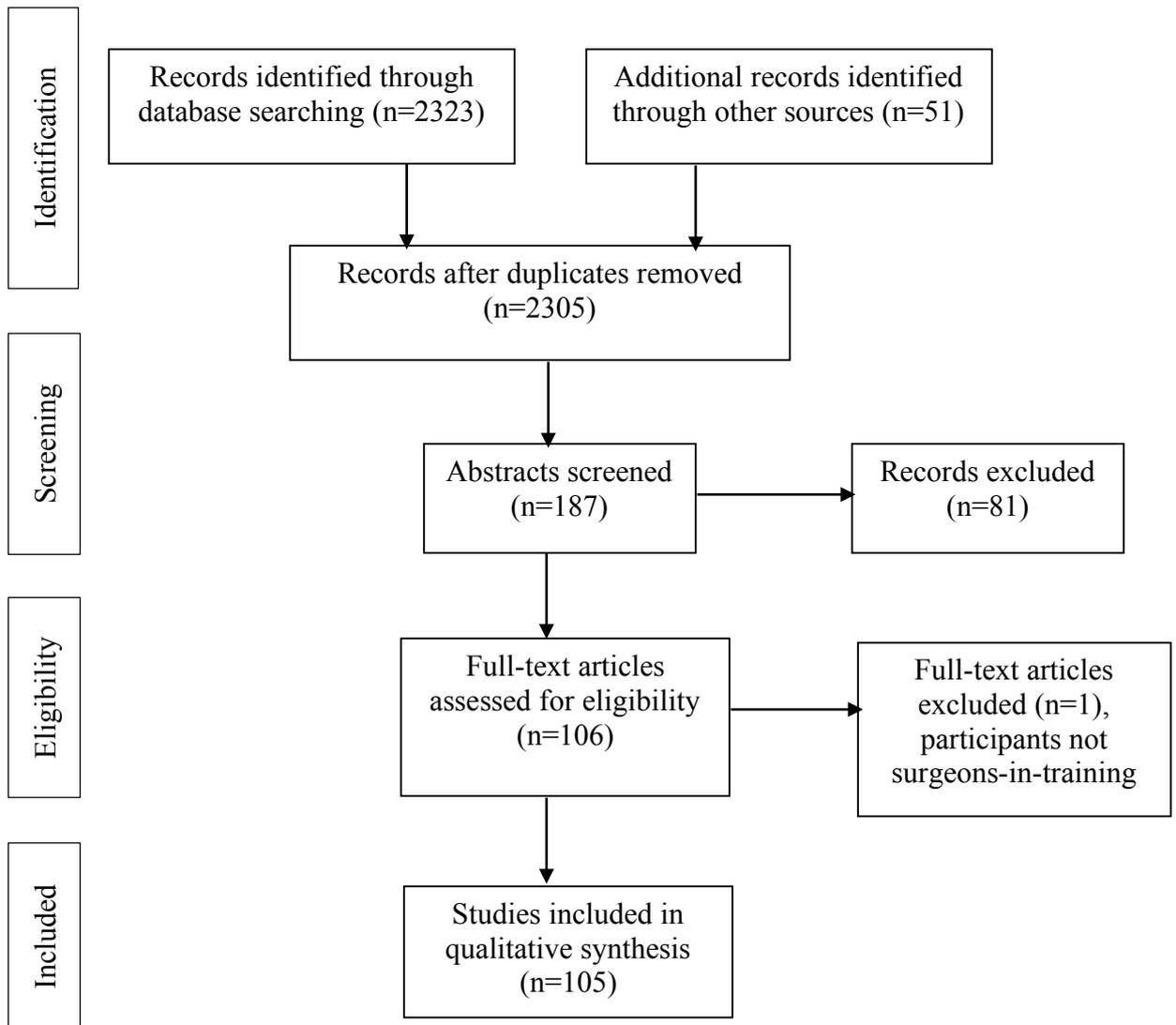


Figure 20. PRISMA diagram

5.4.2 Study aims, setting, assessment format and participants

The studies were broadly split into three categories; 1) studies measuring the impact of a simulation training intervention (26 studies^{255 256 295 296 298 306-326}), 2) studies assessing the construct validity of a simulator designed for training surgeons (42 studies^{274 290 292-294 327-363}) and 3) studies validating an assessment tool (33 studies^{264 364-398}) (tables 8 and 9, column 1).

Of the 106 studies included in the review, 60% assessed arthroscopic skill; 34 knee^{256 290 294 296 298 306 308 312 314 315 327 328 330 331 333 334 338 339 345 346 353 364 365 367-373}, 24

shoulder^{290 295 307 311 314 316 317 329 332 336 337 340 344 346 353 366 367 371 372 378 380-382 399}, four hip^{335 341 342 375}, one ankle³¹³, and six studies assessed basic generic arthroscopic skills^{292 309 310 343 347 374}. The majority of these concerned diagnostic arthroscopy (70%), the procedural arthroscopic skills assessed included arthroscopic Bankart repair (n = 3), rotator cuff repair (n=1), labral repair (n=1), meniscal repair (n=2), anterior cruciate graft preparation (n=2) and insertion (n=1) (shown in table 8 column 5).

There were 42 studies that assessed open surgical procedures. These are shown in table 9, column 5. The open procedures that were assessed in these studies were; DHS (n=4), cannulated hip pinning (n=2) and hemiarthroplasty (n=1) for fractured neck of femur, spinal pedicle screw placement (n=6), open surgical approaches to the shoulder (n=1), hand trauma skills; nail-bed repair, z-plasty, metacarpal fracture fixation, tendon repair (all n=1) and various fracture ORIF; forearm (n= 7), ankle (n=2), tibia (n=1) and complex articular fractures (n=1). Elective hand procedures including trigger finger release (n=1) and carpal tunnel decompression (n=3) and elective hip (n=1) and knee (n=1) arthroplasty were also assessed.

The majority (85 studies) assessed skills in the simulated setting, 10 studies assessed skills in the live operating theatre and 10 studies assessed skills in both the simulated and live operating theatre. Overall, 2088 orthopaedic resident participants were involved in the studies, with experience level ranging from PGY1-10.

5.4.3 Assessment format

Assessment format of the included studies varied considerably, and many studies used more than one method (tables 8 and 9, column 3). Fifty-nine studies assessed performance using live observation and 50 used post-hoc analysis of video footage by experts. Simulator derived metrics were used in 72 studies. Final product analysis by expert assessors was used for 3 studies, and biomechanical testing of the final product was used in 7 studies.

5.4.4 Assessment tools or metrics

There was a large variety of assessment tools used in the review studies (table 10). Objective assessment of technical skills were widely used, and took many forms including; task specific checklists (20 studies), global rating scales (21 studies) and novel objective skills assessment tools in both arthroscopy (22 studies) and open surgery (6 studies). Procedure specific rating scales were used in both arthroscopic (7 studies) and open procedures (6 studies). Individual procedural metrics such as final product analysis and procedure time were used in 68 studies. Movement analysis using simulator derived metrics such as hand movements, gaze tracking, hand position checking and instrument speed and path length were used in 32 studies. Psychomotor testing using commercial dexterity tests were used in 5 studies. Subjective assessment measures were used in 4 studies. Traditional style assessments such as written examinations were used in 5 studies.

5.4.5 Quality Assessment

Each assessment tool was appraised for utility against Van der Vleuten's utility index (results in table 10, columns 5-11). There was a wide spread of utility characteristics amongst the different tools, and the degree of observed heterogeneity precludes any kind of formal analysis. The strengths and limitations of the respective tools, and distribution of evidence across the various domains of the utility index are presented in table 10 columns 3 and 4.

Table 8. Studies assessing arthroscopic performance

Author	Study Aim	Assessment format	Participants and skills assessed	Tools or metrics	Results	Findings relating to assessment tool(s)	
STUDIES IN SIMULATED SETTING							
An et al. ³⁰⁶ , 2018	To determine the effect of simulation training on gaze fixation strategies	SDM	16 PGY?	Diagnostic Knee Arthroscopy	Eye movements, PT	Correlation seen between PT and gaze fixation strategy with repeated instruction	Content, construct and criterion validity of gaze tracking is demonstrated
Angelo et al. ⁴⁰⁰ , 2015	To assess construct validity of the tool on a cadaveric shoulder and establish a proficiency benchmark for arthroscopic Bankart repair	Expert assessor review of video footage	12 PGY 4-5	Arthroscopic Bankart repair	TSC	IRR = 0.92, novice surgeons made 50% more errors and demonstrated increased performance variability and procedure time	Construct validity is demonstrated in the cadaveric model. A proficiency benchmark can be established as the mean performance of the expert group
Angelo et al. ³⁰⁷ , 2015	To compare 3 training protocols for learning to perform an ABR	Expert assessor review of video footage	44 PGY 4-5	3-anchor ABR	Bankart procedure metric tool	A proficiency-based progression training curriculum and protocol coupled with the use of a shoulder model simulator produces a superior arthroscopic Bankart skill set	Bankart procedure metric tool appears construct valid in this setting
Alvand et al. ³⁶⁴ , 2012	To assess the ability of novel visual parameters to objectively discriminate between various levels of arthroscopic experience	Live observation, SDM	15 PGY?	Diagnostic Knee Arthroscopy	HMA, hand position checking, instrument loss, triangulation time	Significant difference in performance between the three groups was seen with visual parameters, GRS and motion analysis	Visual parameters are construct and criterion valid in this setting

Alvand et al. ³⁶⁵ , 2013	Determine if a modified GRS can be used to assess the learning curve during simulated arthroscopic knee meniscal repair	Expert assessor review of video footage and SDM	19 PGY2 -4	Arthroscopic knee meniscal repair	GRS, BAKSS, HMA	Moderate correlation observed between GRS and HMA parameters	Arthroscopic modified GRS can be used for modelling skill development in knee meniscal repair
Bayona et al. ³⁶⁶ , 2014	To assess the validity of the IGARS using a VR shoulder simulator	Expert assessor review of video footage	39 PGY?	Diagnostic shoulder arthroscopy	IGARS	IGARS demonstrated high internal consistency (Cronbach's alpha =0.92). IGARS can distinguish between different levels of experience on a VR simulator	IGARS demonstrates face, content and construct validity. It has high internal consistency and excellent IRR
Bhattacharya et al. ³⁰⁸ , 2017	Evaluate the effectiveness of cognitive task analysis for training in diagnostic knee arthroscopy	Expert assessor review of video footage	16 PGY?	Diagnostic knee arthroscopy	ASSET	The cognitive task analysis showed improved technical skill compared to the control group as measured by the ASSET	ASSET is construct valid in this setting
Brusalis et al. ³²⁷ , 2017	Evaluate a low-fidelity simulation model for ACL graft preparation	Live observation	10 PGY 1-4	ACL graft preparation	Error-focussed scoring	The simulator trained group performed significantly better than controls in overall performance, critical steps and errors	Assessment tool was construct valid in this model
Camp et al. ²⁵⁶ , 2016	Compare the impact of cadaveric simulation training compared to VR training in diagnostic knee arthroscopy	Expert assessor review of video footage	45 PGY 1-5	Diagnostic knee arthroscopy	PT, ASSET	PT and ASSET scores improved in both the cadaveric and VR groups as compared to the controls	ASSET and PT are construct and concurrent valid in this setting
Cannon et al. ²⁹⁴ , 2014	To assess the construct validity of a VR arthroscopy simulator	SDM	6 PGY 1, 6	Diagnostic knee arthroscopy	PT	The mean PT correlated with experience, and was significantly different between PGY 1 and both	Construct validity is demonstrated. Educational benefit is seen in repetitious practice.

			PGY 5			PGY 5 and expert groups. The difference between PGY 1 and 5 was approaching significance (0.055). The level of completeness was NS suggesting all groups achieved a level of consistency in their performance of the procedure. Times improved across 3 trial repetitions in all groups	
Cetinkaya et al. ³⁰⁹ , 2017	Evaluate the impact of a simulation course on psychomotor skill performance	Live observation	70 PGY?	Basic arthroscopy skills	Psychomotor test metrics	Scores improved significantly for time to completion and error rate for all psychomotor tests post-course	Construct validity of psychometric testing tools for measuring basic arthroscopic skills is demonstrated
Chong et al. ³¹⁰ , 2016	Define the early learning of arthroscopic knot tying	Biomechanical testing	3 PGY 1, 3 PGY 3	Arthroscopic knot tying	FPA, PT	Scores improved in the inexperienced group across 3 stages of simulator training	Construct and concurrent validity of knot tensile strength testing demonstrated
Colaco et al. ²⁹² , 2017	To assess the construct validity of three skill deconstructed models	Live observation	12 PGY 1-10	Basic arthroscopy skills	PT, Hand position checking	Average PT and hand position checking frequency correlated inversely with experience level	Construct validity of simulator and educational impact demonstrated

Coughlin et al. ²⁹³ , 2015	To assess construct validity of the simulator model using an aggregate assigned score for the six component tasks	Live observation, expert assessor review of video footage	12 PGY 2-3, 16 PGY 4-5	Basic arthroscopy skills	Aggregate arthroscopic skills score	Construct validity demonstrated by significant improvement in scores by increasing levels of training between all groups. The model was highly reliable	Construct validity and reliability demonstrated
Dwyer et al. ³²⁸ , 2015	To validate dry knee simulator model for assessing performance of ACLR	Live observation	29 PGR 1-5	ACLR	ASSET	Internal reliability using the total ASSET score was very high (>0.9). Construct validity demonstrated in significant observed score differences by level (p<0.05)	ASSET is reliable and construct valid
Dwyer et al. ³⁶⁷ , 2016	To determine if OSATS is valid for assessing residents performance of sports surgery procedures in a competency-based model	Expert assessor review of video footage	27 PGY 1-5	Partial meniscectomy, Whipstitch hamstring graft, ACL femoral tunnel drilling, labral suture passing, insertion of rotator cuff anchors, passing rotator cuff sutures, tie sliding arthroscopic knot	ASSET, OSATS TSC and GRS	A significant difference by PGY was seen for the overall GRS, total ASSET score and total checklist score, as well as for each procedure (p<0.001)	Construct and concurrent validity demonstrated
Dwyer et al. ³²⁹ , 2017	To evaluate the use of dry models to assess performance of arthroscopic rotator cuff repair (RCR) and labral repair (LR) on a dry model	Live observation, expert assessor review of video footage	LR: 39 PGY 1-5, RCR: 35	Arthroscopic RCR and LR	TSC, ASSET non-OSATS GRS	Internal consistency and IRR using total ASSET score was high (>0.9). Construct validity of the model was demonstrated	ASSET demonstrated internal consistency, IRR, construct validity and concurrent validity

			PGY 1-5				
Elliot et al. ³⁶⁸ , 2012	To develop a scoring system to evaluate individual proficiency at diagnostic knee arthroscopy	Live observation (remote for blinding)	20 PGY 1-5	Diagnostic knee arthroscopy	PT, Arthroscopic Skills Assessment form	Statistically significant differences in total score, procedure time and number of missed structures between participants grouped by experience level	The tool is face, content and construct valid for assessing diagnostic knee arthroscopy performance in the simulated setting
Escoto et al. ³³⁰ , 2012	To evaluate the construct validity of a force-sensing simulator for knee arthroscopy skill assessment	SDM	10 PGY?	Diagnostic and therapeutic knee arthroscopy	PT, instrument collision force	Performance difference between experts and novices was seen in 9 of 14 tasks. The difference in force applied to the femur is statistically significant for the two oscillating shaving tasks with the novices applying more force than the experts on average	Force applied to joint structures has construct validity in assessing arthroscopic skill level
Fucentese et al. ³³¹ , 2015	To determine the face and construct validity of a new VR simulator for therapeutic and diagnostic knee arthroscopy	SDM	33 PGY?	Diagnostic and therapeutic knee arthroscopy	PT, instrument path length	Face validity and educational impact of simulator was good. Experts performed better than novices in all domains however performance difference between the three groups individually did not reach significance	SDM not construct valid in this model

Garfield Roberts et al. ³⁵³ , 2017	To assess the face and construct validity of a passive haptic VR simulator	SDM and expert assessor review of video footage	25 PGY? Ω	Diagnostic and therapeutic knee and shoulder arthroscopy	PT, path length, novel SDM: scratching score, % of normal meniscus removed and % of lesion removed. ASSET GRS	ASSET GRS score, PT and path length demonstrated construct validity. Good IRR with ASSET. Mixed construct validity evidence for the novel metrics	ASSET and established SDM metrics showed good construct validity
Gomoll et al. ³³² , 2007	Test the construct validity of a shoulder arthroscopy simulator	SDM	25 PGY2 -5	Diagnostic shoulder arthroscopy	PT, Instrument path length, collisions	A significant association was observed between all tested parameters and level of surgical experience	SDM are construct valid in this model
Gomoll et al. ³¹¹ , 2008	To assess the skill retention/improvement 3 years after arthroscopic simulator training	SDM	10 PGY?	Diagnostic shoulder arthroscopy	PT, instrument path length, collisions and injuries	Subjects improved significantly across all 4 parameters at 3 year retest	PT and path length demonstrated construct validity. Collisions and injuries did not.
Howells et al. ⁴⁰¹ , 2008	Test the construct validity of a motion analysis system for assessing performance of simple arthroscopic tasks	SDM	20 PGY?	Diagnostic and therapeutic knee arthroscopy	HMA	Significant performance differences seen between surgeons and non-surgeons (p<0.0001) and between senior and junior surgeons (p<0.05) – trend towards decreased PT and improved economy of movement with increasing arthroscopic experience	HMA has construct validity in this setting

Insel et al. ³³⁴ , 2009	To develop and validate an objective model for assessing basic arthroscopic proficiency	Live observation	59 PGY 1-5	Diagnostic knee arthroscopy	BAKSS	Strong correlation between GRS scores, YIT and number of previous arthroscopies performed ($r=0.88$, $p<0.01$). TSC scores were moderately correlated with YIT ($r=0.73$, $p<0.01$) and number of previous arthroscopies ($r=0.64$, $p<0.01$)	The BAKSS tool is construct valid for assessing proficiency in diagnostic knee arthroscopy
Jackson et al. ³¹² , 2012	Demonstrate learning curve for meniscal repair and determine impact of task repetition on retention of this skill	Live observation SDM	19 PGY?	Arthroscopic meniscal repair	HMA	All subjects demonstrated a clear learning curve during the initial learning phase. There was no loss of skill seen after a 6-month break on task repetition	HMA can be used to model learning curves in arthroscopic meniscal repair
Khanduja et al. ³³⁵ , 2017	To test the construct validity of the hip diagnostics module of a virtual reality hip arthroscopy simulator	SDM	10 PGY 1-6	Diagnostic hip arthroscopy	PT, error rating	Increased experience in hip arthroscopy was reflected by significantly better performance on the simulator across 2 tasks	Construct validity of SDM demonstrated
Koehler et al. ³⁷⁰ , 2013	To evaluate the content and concurrent validity and reliability of the ASSET for performance of diagnostic knee arthroscopy	Expert assessor review of video footage	28 PGY 1-5	Diagnostic knee arthroscopy	ASSET	Significant score differences between novice, intermediate and advanced groups was seen. Scores were strongly correlated between raters ($r=0.91$, $p<0.01$) and for sequential procedures by each surgeon ($r=0.79$, $p<0.01$)	Construct and concurrent validity, inter-rater reliability and test-re-test reliability were demonstrated for diagnostic knee arthroscopy in the cadaveric model

Koehler et al. ³⁶⁹ , 2013	To test the validity and reliability of the ASSET as a pass-fail examination of arthroscopic skill	Expert assessor review of video footage	28 PGY?	Diagnostic knee arthroscopy	ASSET	Raters agreed on pass-fail rating for 56/60 videos (ICC=0.83). Logging >80 arthroscopic cases or performing more than 35 knee arthroscopies was predictive of passing	ASSET is valid and reliable as a pass-fail examination of diagnostic arthroscopy of the knee in the simulation setting
Martin et al. ³³⁶ , 2011	Evaluate the correlation between timed task performance in an arthroscopic shoulder simulator and timed task performance in a cadaveric shoulder arthroscopy model	Live observation, SDM	15 PGY?	Diagnostic shoulder arthroscopy	PT	PT on the simulator was strongly correlated with PT on the cadaveric model (r=0.736, p<0.001), and experience predicted performance in both models (p=0.016)	PT is a construct valid measure of arthroscopic skill in VR and cadaveric simulation settings
Martin et al. ³³⁷ , 2012	To evaluate the correlation between timed task performance on an arthroscopic shoulder simulator and resident experience	SDM	27 PGY 1	Diagnostic shoulder arthroscopy	PT	PT correlates with experience. For every PGY increment, there was a 16 second improvement in the time required to complete the simulator task (p<0.005)	PT as measured on the simulator is a construct valid measure of arthroscopic skill
Martin et al. ³¹³ , 2015	To determine if low-fidelity arthroscopic simulation training improves basic ankle arthroscopy performance	Expert assessor review of video footage	29 PGY 1-5	Diagnostic ankle arthroscopy	ASSET	Simulation group outperformed the control group in total ASSET and checklist scores (p<0.001)	ASSET is construct valid
Martin et al. ³⁹⁹ , 2016	To evaluate the correlation between timed task performance on a VR shoulder arthroscopy simulator and participation in an expert arthroscopy course	SDM	99 PGY?	Diagnostic shoulder arthroscopy	PT, path length	Significant improvements in PT and path length (camera and probe tip) were seen after training	PT and path length are construct valid in this VR model

McCarthy et al. ³³⁸ , 1999	To evaluate the construct validity of the Sheffield Knee Arthroscopy Trainer	SDM	6 PGY?	Diagnostic knee arthroscopy	PT, probe collisions	Experienced surgeons performed best with fewer instrument collisions	PT and error counting is construct valid measure of arthroscopic skill
McCarthy et al. ³³⁹ , 2006	To evaluate the Sheffield Knee Arthroscopy Training System	SDM	13 PGY?	Diagnostic knee arthroscopy	PT, path length, collisions	Experienced surgeons performed better with shorter procedure time, probe path length and found more pathology. After SKATS training, novices showed improvements across all domains	SDM construct valid in this simulator model
Middleton et al. ³⁷¹ , 2016	To compare three GRS tools for the assessment of simulated arthroscopic skills	SDM, expert assessor review of video footage	21 PGY?	Knee: diagnostic arthroscopy, basic triangulation and medial meniscectomy. Shoulder: diagnostic arthroscopy, basic triangulation task, advanced triangulation task	ASSET, BAKSSS, IGARS	All GRS demonstrated construct validity with significant differences between each skill level and arthroscopic task. IRR was high for all. Correlation with time taken and path length was significant for all	No single GRS tool demonstrated superiority
Nwachukwu et al. ³⁷² , 2016	To test the ability of a procedure specific checklist to detect performance improvement over time	Live assessment	21 PGY 3	Diagnostic knee arthroscopy, diagnostic shoulder arthroscopy	Diagnostic Shoulder Arthroscopy Checklist, Diagnostic Knee Arthroscopy Checklist	Mean time to checklist (procedure) completion improved pre- and post-intervention (6/52)	The checklists are construct valid to measure operative efficiency. They do not measure quality of performance.

Olson et al. ³⁷³ , 2013	To test the construct validity and reliability of a modified BAKSSS (GRS only) using mixed-level assessors	Live assessment	23 PGY1 -5	Diagnostic knee arthroscopy	BAKSSS GRS	The modified BAKSSS demonstrated construct validity with junior residents achieving lower scores (mean 20 vs mean 30 for senior residents).	The BAKSS GRS showed construct validity and fair inter-rater reliability
Pedowitz et al. ³⁴⁰ , 2002	To evaluate the construct validity of a novel shoulder VR simulator	SDM	22 PGY?	Diagnostic shoulder arthroscopy	Instrument path length/ratio, error scoring	Test time and path ratio differed significantly as a function of surgical experience	Limited construct validity of SDM
Pedowitz et al. ³⁷⁴ , 2015	To a) assess a new biomechanical assessment of arthroscopic knots and to b) establish proficiency benchmarks using the fundamentals of arthroscopy trainer	Biomechanical testing	44 PGY4 -5	Arthroscopic knot tying	FPA	Performance was inconsistent between experience levels, and failure rate could not predict experience level	The results suggest that FPA of arthroscopic knot quality is not construct valid, or that there might be performance heterogeneity amongst even the most experienced arthroscopic surgeon
Phillips et al. ³⁷⁵ , 2017	Evaluate the use of a TSC, ASSET, PT and GRS for assessing performance of arthroscopic hip labral repair in a dry model	Expert assessor review of video footage	37 PGY?	Arthroscopic hip labral repair	TSC, ASSET, novel GRS, PT	Dry model to assess the performance of arthroscopic labral hip repair by residents is both reliable and valid	ASSET demonstrated IC, IRR, construct validity and concurrent validity
Pollard et al. ³⁴¹ , 2012	To compare learning curves in diagnostic hip arthroscopy on a low fidelity simulator, in lateral and supine positions	SDM	20 PGY 1-4	Diagnostic hip arthroscopy	PT, HMA	Both groups demonstrated learning in all parameters ($p < 0.001$). Junior participants achieved performance parity with senior participants after 12 attempts	Construct validity of SDM demonstrated
Price et al. ³⁷⁶ , 2015	Evaluate the number of arthroscopies needed to achieve consultant level performance	Expert assessor review of video footage, SDM	28 PGY 1-8	Diagnostic knee arthroscopy	BAKSSS GRS, HMA, PT	There was significant improvement in performance with increasing experience ($p < 0.05$)	BAKSSS, PT and HMA is construct and concurrent valid

Rahm et al. ³⁴² , 2016	To assess face and construct validity of a VR based model for shoulder arthroscopy	SDM	25 PGY?	Diagnostic hip arthroscopy	PT, instrument path length	Simulator positively rated as educationally valuable. Experts were significantly faster and demonstrated better economy of movement	Construct validity of metrics demonstrated
Rahm et al. ³¹⁴ , 2018	Test a standardised, competency based training protocol on a VR arthroscopy simulator	Expert assessor review of video footage, SDM	20 PGY 1-5	Diagnostic knee and shoulder arthroscopy	ASSET, PT, instrument path length	The residents' performance significantly improved post-training in ASSET and SDM parameters. Expert ASSET score was significantly higher than the residents post-training score.	ASSET was construct valid. SDM showed a mixed validity picture; for shoulder it was construct valid, but not for knee
Rebolledo et al. ²⁹⁵ , 2015	To compare performance of arthroscopic simulator-trained and didactic-trained residents in diagnostic knee arthroscopy	Live observation	14 PGY 1-2	Diagnostic shoulder arthroscopy	PT, IGI	Participants trained on the simulator outperformed the didactic trained group on the cadaveric model test by PT (-35%, p=0.02) and IGI (-35%, p=0.01)	Educational value of simulator demonstrated. IGI appears to be construct and concurrent valid but not formally evaluated here
Rose et al. ³⁴³ , 2015	To assess the construct validity of three skill-deconstructed VR models	SDM	10 PGY?	Basic arthroscopy skills	PT, HMA	2 assessments (Steady and Probe and Track a Moving Target) demonstrated construct validity. 1 assessment type did not (Steady and Telescope)	VR task deconstruction could be used for assessment and development of early arthroscopic skills
SladeShantz et al. ³⁷⁷ , 2013	To determine whether a global assessment of arthroscopic skills was valid for blinded assessment of cadaveric diagnostic arthroscopy	Expert assessor review of video footage	13 PGY 1-3	Diagnostic knee arthroscopy	OAAS	The agreement between global assessment scores was strong (ICC=0.80). Internal consistency was excellent (Cronbach's alpha=0.97) and test re-test reliability was strong. The test is construct valid and able to discriminate	The OOAS is reliable and construct valid in this setting

						between various levels of training (p<0.0001)	
Srivastava et al. ³⁴⁴ , 2004	To evaluate the construct validity of a novel shoulder simulator	SDM	21 PGY?	Diagnostic shoulder arthroscopy	PT, collisions, errors	The expert group performed the task more quickly (p=0.013) and more accurately (p=0.002)	Construct validity of procedural accuracy demonstrated. PT not construct valid in this model
Tuijthof et al. ²⁹⁰ , 2010	To evaluate face and construct validity of a knee arthroscopy simulator	SDM	8 PGY?	Diagnostic knee arthroscopy, diagnostic shoulder arthroscopy	PT	The experienced (consultant) group were more time efficient than the residents in task completion for each repetition (p<0.05). The participants reported overall good face validity of the simulator	Face and construct validity of the simulator is demonstrated
Tashiro et al. ³⁴⁵ , 2008	To test the construct validity of a knee arthroscopy simulator with force evaluation	SDM	24 PGY?	Diagnostic and therapeutic knee arthroscopy	PT, instrument path length, instrument forces	The experienced group performed both tasks more efficiently and competently than the less experienced groups. Path length was shorter, velocity faster and forces applied lower	SDM construct valid in this model
Tofte et al. ³⁴⁶ , 2017	To validate a VR arthroscopy simulator	SDM	35 PGY 1-6	Diagnostic knee and shoulder arthroscopy	PT, instrument path length	Significant correlations between experience level and performance as measured by PT and instrument path length	PT and SDM construct valid in this model

Wong et al. ³⁴⁷ , 2015	To test the construct validity of the Arthroscopic Knot Trainer	Expert assessor review of video footage	21 PGY 1-3, 11 PGY 4-5	Arthroscopic knot tying	PT, number of knots tied in pre-specified time limit	The simulator could discriminate performance by level of experience. Positive qualitative feedback on educational value	PT construct valid in this model
STUDIES IN LIVE THEATRE SETTING							
Cannon et al. ³¹⁵ , 2014	To assess the transfer validity of skills learnt on a VR simulator to the operating room	Expert assessor review of video footage	54 PGY 3	Diagnostic knee arthroscopy	TSC and GRS	The simulator trained group performed significantly better as measured by the TSC (p=0.031), including probing skills (p=0.016) but not visualisation skills (p=0.34), compared to the control group (the TSC weighted probing skills double the weight of visualisation skills). The GRS failed to reach significance, probably because of an extreme outlier	The TSC has a mixed validity picture, and the weightings of probing skills vs visualisation skills may need adjusting. GRS not construct valid in this setting
Dunn et al. ³¹⁶ , 2015	To test the impact of a simulation training intervention on a) the initial skill improvement, and b) retention of skill longitudinally in performance of diagnostic shoulder arthroscopy	Live observation	17 PGY 1-5	Diagnostic shoulder arthroscopy	ASSET	No performance difference between simulator trained and untrained at baseline, the simulator trained group improved as compared to baseline in mean ASSET (p=0.023) and PT (p=0.01). The training effect was lost by 12 months follow-up.	ASSET is construct and concurrent valid in the live theatre setting

Gallagher et al. ³⁷⁸ , 2018	To determine the IRR of a TSC	Expert assessor review of video footage	44 PGY 4-5	Arthroscopic Bankart Repair	TSC	Mean IRR = 0.93 (range 0.84-0.99)	High IRR demonstrated, tool suitable for high stakes assessment
Garfield Roberts et al. ²⁹⁸ , 2019	Investigate transfer validity of simulation training using elbow-worn motion sensors	SDM	30 PGY 2-3	Diagnostic Knee Arthroscopy	Elbow motion analysis, PT	The intervention group outperformed the control group in all metrics	Intra-operative elbow motion analysis metrics are construct valid in the live theatre setting
Hodgins et al. ³⁷⁹ , 2014	To describe the learning curve for diagnostic knee arthroscopy	Live observation	20 PGY 1-5	Diagnostic Knee Arthroscopy	BAKSSS, GRS, LC-CUSUM	Competency as assessed by the TSC was achieved by 40% of trainees after a median of 16 procedures and by the GS for 1 trainee	LC-CUSUM is an effective method to evaluate procedural competence in arthroscopic training and can provide objective feedback and benchmarks in the learning phase
Howells et al. ²⁹⁶ , 2008	To evaluate the transfer validity of arthroscopic skills from simulator training to the operating theatre	SDM, live observation	20 PGY 1-2	Diagnostic Knee Arthroscopy	HMA, OSATS GRS and PBA	Simulator trained group performed significantly better than untrained group using PBA and OSATS GRS demonstrating the transfer of skills from simulator to live theatre	HMA has construct and criterion validity in this setting
Hoyle et al. ³⁸⁰ , 2012	Develop and validate a new GRS for shoulder arthroscopy	Expert assessor review of video footage	10 PGY?	Diagnostic shoulder arthroscopy	GRSSA	Good construct validity, mixed reliability	GRSSA is construct valid for assessing performance in diagnostic shoulder arthroscopy, and can be used remotely from the procedure
Koehler et al. ³⁸¹ , 2015	To determine the validity and reliability of using ASSET to assess arthroscopic skill in the operating room	Expert assessor review of video footage	8 PGY 3-5	Diagnostic shoulder and knee arthroscopy	ASSET	The senior group achieved significantly higher mean ASSET scores compared to junior group for both procedures. ICC for total scores was good (knee =	ASSET is feasible, reliable and construct valid for assessing diagnostic arthroscopy in live theatre

						0.81, shoulder = 0.84). Raters concurred on pass-failure evaluation in 89% of procedures	
Talbot et al. ³⁸² , 2015	To assess the reliability and validity of the shoulder OPAT when performing diagnostic shoulder arthroscopy	Live observation and expert assessor review of video footage (for IRR assessment)	6 PGY 3-10	Diagnostic shoulder arthroscopy	Shoulder OPAT	Internal consistency = 0.77, IRR=0.6, IRR=0.82. Face, content, construct and concurrent validities were demonstrated.	Shoulder OPAT fulfils several aspects of reliability and validity and is perceived as superior to PBA
Waterman et al. ³¹⁷ , 2016	To assess the training impact of a shoulder arthroscopy simulator model	SDM, Expert assessor review of video footage	22 PGY?	Diagnostic shoulder arthroscopy	ASSET, PT instrument path length	The simulator trained group showed improvement in ASSET score in the live theatre environment. There was no significant improvement in PT seen in the live environment	ASSET is construct valid in this setting.

Ω - Number includes medical students

ACLR = Anterior Cruciate Ligament Reconstruction
ASSET = Arthroscopic Surgical Skill Evaluation Tool
BAKSSS = Basic Arthroscopic Knee Skills Scoring System
COR = Cut Out Rate
FPA = Final Product Analysis
GRS = Global Rating Scale
GRSSA = Global Rating Scale for Shoulder Arthroscopy
HMA = Hand Motion Analysis
ICC = Intraclass Correlation Coefficient
IGI = Injury Grading Index
IGARS = Imperial Global Arthroscopy Rating Scale
IRR = Inter-rater reliability
LC-CUSUM = Cumulative Summation Test for Learning Curve
LR = Labral repair

NS = Not Specified
NOF = Neck of Femur
OAAS = Objective Assessment of Arthroscopic Skills
OPAT = Objective practical assessment tool
PGY = Postgraduate Year
PT = Procedure Time
RCR = Rotator cuff repair
SDM = Simulator Derived Metrics
TAD = Tip Apex Distance
TSC = Task Specific Checklist
VR = Virtual Reality
YIT = Year in training

Table 9. Studies assessing open surgical performance

Author	Study Aim	Assessment format	Resident Participants	Skills Assessed	Tools or metrics	Results	Findings relating to assessment tool(s)
STUDIES IN SIMULATED SETTING							
Akhtar et al. ³⁴⁸ , 2015	To assess the construct validity of a VR trauma simulator for performing DHS fixation of trochanteric fracture	SDM	10 PGY1-4, 10 PGY5-12	DHS fixation of trochanteric fracture	FPA, fluoroscopy time	Statistically significant differences seen across 5 of 6 performance indices (excluding total procedure time) which correlated with frequency of exposure of operating	Construct validity of FPA demonstrated in the simulated model
Anderson et al. ³⁸³ , 2016	To determine if OSATS predicts surgical quality of the procedure	Live observation, biomechanical testing, FPA	21 PGY 1-2 (A) + 30 PGY 2-5 (B)	A) Tibial plafond fracture reduction B) Distal radius ORIF	OSATS, FPA	OSATS did not correlate with the articular reduction quality (A) or the integrity of the mechanical fixation (B)	OSATS do not effectively assess the quality of the surgical result
Aoude et al. ³⁴⁹ , 2016	To assess the utility of computer-assisted surgery in pedicle screw placement	Expert assessor review of CT images	24 PGY 1-4	Pedicle screw placement	FPA	Experience level did not predict screw placement accuracy	FPA not construct valid in this setting
Backstein et al. ³⁸⁴ , 2004	To compare the effect of three types of feedback on resident post-test performance; control group (no feedback), video and self-review and expert feedback	Expert assessor review of video footage	29 PGY 1-5	Plating of long bone fracture, TBW olecranon, Z-plasty	Video assessment with feedback	GRS scores across the 3 groups and tasks were not significantly different	Video feedback assessment failed to demonstrate an improvement in technical skills

Bergeson et al. ²⁶⁴ , 2008	To evaluate the early learning curve of pedicle screw placement	Live observation	3 PGY1 and 3	Thoracic pedicle screw placement	FPA	Acceptable placement accuracy levels were achieved by the fourth attempt	Construct validity and educational impact of FPA is demonstrated
Bernard et al. ³⁸⁵ , 2016	Compare the reliability and validity of OSATS checklist, GRS score and subjective pass/fail to assess resident operative skill in shoulder surgery	Live observation	23 PGY 1-5	Open surgical approaches to the shoulder (deltopectoral, lateral deltoid-splitting and posterior)	OSATS checklist, GRS, pass-fail assessment	Concurrent validity shown between OSATS and GRS for the 3 shoulder approaches. OSATS has superior reliability compared with GRS and pass/fail	OSATS checklists are a valid and reliable assessment of technical skills across 3 open surgical shoulder approaches
Blyth et al. ³⁵⁰ , 2008	To test the construct validity of a computer-based VR simulator	SDM	6 PGY1-2, 6 PGY3-5	DHS	PT, composite score derived from simulator metrics	Accuracy, number of radiographs and speed were significantly different between experience levels, critical error rate was inversely associated with experience level. PT was not predictive of experience	Radiograph use, speed and critical error rate demonstrate construct validity. PT did not in this model
Burns et al. ³⁸⁶ , 2017	To evaluate the educational value of low fidelity ORIF simulation and objective measures of performance using biomechanical means	Biomechanical testing	16 PGY 1	Ulnar ORIF	FPA	Pre-to post-simulation operative success rates were significantly improved (p<0.001) and were maintained at 3/12 follow-up	Biomechanical construct failure testing is a valid method of measuring operative success in the simulated setting

Butler et al. ³¹⁸ , 2017	Evaluate a training course for interns on closed reduction and pinning of paediatric supracondylar fractures	Live observation and written examination	19 PGY 1-5	Closed reduction and pinning of paediatric supracondylar fractures	TSC and written exam	Post-training there was no difference in MCQ scores and TSC score comparing interns and senior residents (significant difference prior)	Written examination and TSC construct valid in this model
Christian et al. ³⁵¹ , 2018	Assess whether VR simulation platform can distinguish between novice and experienced surgeons when performing percutaneous hip pinning	SDM	17 PGY 2-5	Percutaneous fixation of valgus impacted fractured NOF	FPA, PT, fluoroscopy time	Significant association between performance and experience in 10/15 outcome measures.	FPA and PT construct valid in this model. Fluoroscopy time not construct valid.
Froelich et al. ³⁵² , 2011	Evaluate the construct validity of a haptic VR surgical simulator for assessing performance in centre-centre guidewire insertion during intertrochanteric proximal femoral fracture fixation	SDM	15 PGY1-5	Guidewire insertion for DHS	FPA, PT, fluoroscopy time	No difference seen between groups for procedure time or TAD. Significant difference observed in final wire position on lateral view (p=0.01), wire passes (p=0.03) and fluoroscopy time (p=0.05)	Mixed picture of construct validity although some measures show discriminatory ability. Potential for educational impact
Giurin et al. ³⁵⁴ , 2018	To evaluate a low fidelity simulator for nail bed repair	Live observation	12 PGY?	Nail-bed repair	PT	PT was significantly faster for the experienced group (p=0.01)	PT is construct valid in this model

Gottschalk et al. ²⁵⁵ , 2015	To analyse the training effect of 3D simulation on lateral mass screw placement	Expert assessor review of images	15 PGY 1-6	Lateral mass screw placement	FPA	Subjects in the 3D training groups showed significantly improved drilling trajectories as compared to the controls (p<0.0001)	FPA as measured by aggregate mean difference from ideal has construct validity in this model
Hohn et al. ³⁵⁵ , 2014	To evaluate a novel set of low fidelity bone training models	Live observation	15 PGY 1-5	Basic orthopaedic surgical skills	FPA	The model was feasible. No significant difference in performance between experience levels was seen	FPA was not construct valid in this low fidelity model
LeBlanc et al. ³¹⁹ , 2013	To compare task performance on a novel fracture simulator model compared to a bench-top model	Live observation	22 PGY 1-5	Surgical fixation of ulnar fracture	OSATS, PT	Both simulators distinguished between different experience levels, participants performed significantly better on the virtual simulator compared to bench model in all measures except PT (<0.05)	Construct validity for OSATS demonstrated, not for PT in this model
Leong et al. ³⁵⁶ , 2008	To evaluate the validity of three low cost models of fracture fixation in the assessment of technical skills	HMA, live observation	15 PGY?	DCP application, IM nail insertion, wrist ex-fix application	HMA, OSATS GRS	OSATS score differed significantly between the 3 groups and 3 procedures with high IRR (Cronbach's alpha = 0.88). Motion analysis distinguished between the three groups on the DCP model but a ceiling effect was observed in the IM nail and ex-fix procedures	The DCP cadaveric porcine model has face, content and construct validity. The IM nail and ex-fix models are less sensitive and further work is needed

Lopez et al. ³⁵⁷ , 2015	To assess the construct validity of a cost effective psychomotor assessment tool; the Fundamentals of Orthopaedic Surgery (FORS)	Live observation	58 PGY?	Basic psychomotor surgical tasks	FORS	Stratification by experience level was seen in all 6 tasks	Construct validity demonstrated. Educational benefit seen in repetitious practice
MacEwan et al. ³⁸⁷ , 2016	To compare the O-SCORE to the OSATS in assessing performance of a simulated radius ORIF	Expert assessor review of video footage	19 PGY 1-5	ORIF midshaft radius	O-SCORE, OSATS checklist, GRS	O-SCORE demonstrates accurate and reproducible results compared to current gold standard tools (OSATS/GRS) when used in a randomised blinded fashion	Construct, concurrent validity and reliability of O-SCORE demonstrated
Mayne et al. ³⁵⁸ , 2016	To evaluate a distal radius fracture simulator	Expert assessor review of video footage	20 PGY 1-5	Distal radius ORIF	OSATS, FPA	Significant performance differences between junior and senior residents was seen with OSATS (p<0.001). No difference in FPA was seen. IRR was high for all measures	OSATS is construct valid. FPA is not in this model.
Nousiainen et al. ³²⁰ , 2013	To compare performance change in naïve participants undergoing training in cannulated hip pinning on a simulator, using either computer assisted or conventional	SDM	52 PGY 1/Medical Students	Cannulated hip pinning for fractured neck of femur	FPA, Image intensifier use	All participants improved in hardware placement accuracy after the training (p<0.001) and the skill level was retained at post-testing, retention testing at 4/52 and transfer testing onto the opposite study arm. No significant change in guidewire parallelism	FPA and image intensifier use were construct valid in this setting

	fluoroscopic techniques					was seen. Radiation time and dose varied by intervention but not reported by experience level	
Pedersen et al. ³⁸⁸ , 2014	To develop and validate a tool to assess hip fracture fixation performance and set pass/fail benchmark	SDM	10 PGY 1	3 internal fixation procedures for undisplaced femoral neck fracture	Hip fracture performance score	The combined score could discriminate between experience levels. None of the individual SDM demonstrated discriminatory abilities	Construct validity and reliability of the score demonstrated
Putnam et al. ³⁸⁹ , 2015	Assess whether standardised OITE, OSATS and simulation based computer-animated testing can predict biomechanical construct performance following volar plating of distal radial fractures in cadavers	Live observation, SDM and biomechanical testing	15 PGY 3-4	Volar plating of distal radius fracture	Fracture specific knowledge test, OITE, OSATS, GRS, TSC	No statistically significant correlation seen between performance on biomechanical testing and that of knowledge tests, OITE, GRS or TSC	Traditional written and computer-based testing methods failed to predict which fracture constructs would pass biomechanical testing
Qassemnyar et al. ³⁵⁹ , 2014	Evaluate the face and construct validity of a low-fidelity hand trauma examination model	Live observation	19 PGY 3-4	Hand trauma skills; Z-plasty, metacarpal fracture fixation, tendon repair	OSATS	There was significant performance differences seen between microsurgery experience levels	OSATS was construct valid in this model
Rambani et al. ³⁶⁰ , 2014	To validate a desktop simulator for spinal surgery	Live observation and SDM	12 PGY?	Lumbar pedicle screw insertion	TSC	Improved performance in simulator trained group	TSC appears to show construct validity in this model

Ruder et al. ³²¹ , 2017	Evaluate the effectiveness of a training session in orthopaedic drilling technique	Live observation	5 PGY 1	Bone drilling	FPA	Plunge depth was significantly reduced in both groups post-training (p<0.05)	FPA was construct valid in this model
Sonnadara et al. ³²² , 2011	To assess whether an intensive lab based skills course at the start of orthopedic residency is effective for teaching core skills	Live observation	22 PGY 1	Basic orthopedic surgical skills	OSATS	Residents in the simulation trained group performed better in in both checklist and GRS components of OSATS in post-training assessment	OSATS are a construct valid
Sonnadara et al. ³²³ , 2012	To examine retention rates for basic surgical skills taught at the beginning of residency	Live observation	18 PGY1-3	Basic orthopedic skills	OSATS	Mean GRS score for competency based curriculum (intervention) group was 4.3 and maintained 6/12 later. Both intervention and senior control group performed better than the junior resident control group (p<0.001)	OSATS can be used longitudinally to assess skill retention
Sonnadara et al. ³²⁴ , 2013	Compare student-led and instructor-led techniques in improving performance in 4 skills during a surgical boot camp	Live observation	12 PGY 1	Basic orthopaedic skills	OSATS	Checklist and GRS scores were improved in the student led group	OSATS is construct valid
Shi et al. ³⁶¹ , 2018	Assess validity of VR simulator to teach lumbar pedicle screw placement by comparing VR trained and untrained	Expert assessment of final product (scan images)	10 PGY? (procedure naïve)	Lumbar pedicle screw insertion	FPA	VR trained group had significantly lower screw penetration rates, improved penetration distances and placement	FPA in this setting is face, content and construct valid

	resident performance on cadaveric specimens					acceptability compared to the untrained group	
Sugand et al. ³⁶² , 2018	To validate a digital fluoroscopic simulator for guidewire insertion	SDM	26 PGY 2-9	Dynamic Hip Screw guidewire insertion	PT, FPA	The expert group achieved significantly better TAD and COR. PT was not significantly different between experts and novices	TAD and COR had construct validity. PT did not in this model
Tonetti et al. ³²⁵ , 2009	To evaluate the educational value of a path simulator under fluoroscopic guidance in performing percutaneous iliosacral screw insertion	Live observation	23 PGY?	Percutaneous iliosacral screw insertion simulator	Number of radiographs used, iatrogenic index	Significant differences in the number of intra-operative radiographs was seen between experience, prior procedural and technical knowledge subgroups. Iatrogenic index scores were not significantly different between groups	Intra-operative radiograph use has construct validity in this setting. Iatrogenic index failed to show construct validity
Williams et al. ³⁹⁰ , 2017	Test the utility of a battery of 5 motor skills tests (Grooved Pegboard test and 4 novel tests) to evaluate potential orthopaedic residents	Live observation	30 PGY1-4 residents	Closed fracture reduction, drilling, dexterity and visuospatial skills	Psychometric test metrics	Residents performed better than non-residents (p<0.0001) in every exercise but the drilling test	Concurrent validity demonstrated by comparison to Grooved Pegboard test (internal control). Construct validity demonstrated in 3 of 4 tests
Van Heest et al. ³⁹¹ , 2009	To evaluate the reliability and validity of 4 tests for assessing competence in performing CTR;	Live observation	28 PGY 1-6	CTR	TSC, GRS, pass/fail test, knowledge test	Correlation between YIT and knowledge scores (p<0.001) YIT and detailed checklist scores (p=0.002), YIT and GRS scores (p=0.04) and YIT	Knowledge test and OSATS construct valid

	knowledge test, detailed TSC, GRS, pass/fail assessment					and pass/fail ($p < 0.001$). No correlation was seen between YIT and PT.	
Van Heest et al. ³⁹² , 2012	To evaluate OSATS for 3 common upper extremity procedures	Live observation	27 PGY 2-5	TFR, CTR, DRFF	TSC, GRS, PT, pass/fail test	Construct validity between YIT and checklist scores demonstrated for TFR and CTR, between YIT and GRS scores for TFR and DRFF and between YIT and pass/fail for TFR. Criterion validity demonstrated between GRS, checklist scores and pass/fail for all procedures. Reliability poor. Participants rated OSATS as educationally valuable	OSATS construct valid and educationally valuable. Not reliable in this model
Xiang et al. ³²⁶ , 2014	Test the impact of a pre-operative planning simulator on junior surgeons' pedicle screw insertion accuracy	Live observation, SDM	2 PGY?	Pedicle screw insertion	PT, FPA	PT and positional accuracy improved with training (self-controls) and experts outperformed trainees	PT and FPA construct valid here
Yehyawi et al. ³⁶³ , 2013	1. To develop a physical model to improve articular fracture reduction skills. 2. To develop objective assessment methods and 3. To assess the construct	HMA, Live observation	7 PGY 1/2, 5 PGY 4/5	Complex articular fracture fixation	FPA, PT, HMA metrics, GRS	Significant difference seen in cumulative hand distance travelled between junior and senior residents. There was no difference seen in FPA, PT or GRS	Hand movement economy has construct validity in this setting

	validity of the simulation						
STUDIES IN LIVE THEATRE SETTING*							
Beard et al. ³⁹³ , 2011	Compare the acceptability, reliability and validity of three WBA methods of assessing surgical skill	Live observation	Not stated ^c	Not stated	PBA	PBA showed high reliability for assessing same index procedure (G>0.8) Good correlation was seen with PBA scores and specialist training level, years of surgical training and recent index exposure(r=0.31-0.71). Within test correlations were strong between checklist and global ratings for PBA (r=0.73)	PBA showed construct and concurrent validity, reliability and feasibility
Davies et al. ³⁹⁴ , 2018	Develop and validate a new operative assessment tool that addresses current barriers	Live observation	49 PGY 3-10	Not stated	GOSLE	Strong correlation between PBA and PBA ratings (r=0.87, p<0.001). Positive participant feedback on feasibility and educational impact	Psychometric performance of the GOSLE is comparable to the PBA but is preferred by trainees and trainers. High-quality, actionable feedback better supports formative assessment
Gofton et al. ³⁹⁵ , 2012	To pilot and evaluate the validity and reliability of the O-SCORE	Live observation	22 PGY 1-5	ORIF Wrist, ORIF Hip, ORIF ankle, hemiarthroplasty, THA, knee arthroscopy	O-SCORE	Total performance scores were discriminated by year, PGY 1-2 lower than PGY 3 (p<0.001) and	O-SCORE successfully demonstrates construct validity, reliability, feasibility and educational impact

						PGY 3 lower than PGY 4-5 (p<0.02)	across various PGY levels regardless of procedure type
Hawkes et al. ³⁹⁶ , 2017	To evaluate the Fracture Fixation Assessment Tool for assessing the quality of surgical fixation	Expert assessor review of final product (radiographs)	Unknown number of residents performed test cases	Various fracture fixation procedures; ORIF bimalleolar ankle/radius/ulnar/distal humerus/tibial plateau, TBW Olecranon/patella, IM nail tibia, DHS	The Fracture Fixation Assessment Tool	Internal consistency = 0.98, inter-rater reliability = 0.71, CVR = 0.65	Internal consistency, reliability and content validity is demonstrated. May have applications in service evaluation as well as training
Hoffer et al. ³⁹⁷ , 1990	To see if psychomotor tests on entry and exit to specialist training correlate with faculty ranking of surgical skill	Live observation	8 PGY 1	Unspecified	The Crawford small-part dexterity test, steadiness hole test (unsupported), Bennett hand-tool test. Purdue peg board test and Minnesota rate manipulation test	The rankings from the Crawford and steadiness tests had significant correlation to subjective performance ranking by faculty	Psychomotor performance at entry and exit of specialist training appears to predict surgical skill as assessed by subjective measures
Marriott et al. ³⁹⁸ , 2010	Evaluate the validity, reliability and acceptability of the PBA	Live assessment	81 PGY? across 6 specialities (7 in T&O)	Hip replacement and Knee replacement	PBA	Construct validity demonstrated by correlation of PBA scores with measures of surgical training and experience. Reliability of the adjusted total item score and GSS for any given procedure was	PBA demonstrates construct validity, this evidence is limited to 2 T&O procedures performed by 7 trainees

						achieved using 4 and 3 assessor judgements	
--	--	--	--	--	--	--	--

∞ Beard et al study involved multiple surgical specialities, by-speciality data was not presented

* TSC part-validated in live theatre environment, study took place in simulation lab

COR = Cut out rate

CTR = Open carpal tunnel release

CVR = Content validity ratio

DCP = dynamic compression plate

DHS = Dynamic Hip Screw

DRFF = Distal radius fracture fixation

Ex-fix = external fixator

FORS = Fundamentals of Orthopaedic Surgery

GOSLE = Generic Operative Supervised Learning Event

GSS = Global Summary Scale

IM = Intramedullary

OITE = Orthopaedic In-Training Examination

ORIF = Open reduction internal fixation

PBA = Procedure Based Assessment

TBW = Tension Band Wire

TFR = Trigger finger release

THA = Total Hip Arthroplasty

T&O = Trauma & Orthopaedics

Table 10. Utility evidence of the assessment tools used to evaluate surgical competency in T&O

1. TRADITIONAL ASSESSMENTS										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
A) WRITTEN EXAMINATION										
Orthopaedic In Training Examination (OITE) ^{402 403}	Summative examination (US)	Cost-effective. Quantifies knowledge within set parameters. Standardised, objective and efficient for mass testing	Knowledge does not necessarily equate to skill	Y ⁴⁰³	Y ⁴⁰³	Y ⁴⁰² N ³⁸⁹	Y ³⁸⁹ , 403	Y ⁴⁰²	-	-
Web-based cognitive test ³⁹²	Multi-modal 100 point test			Y ³⁹²	-	M ³⁹²	N ³⁹²	-	-	-
Fracture-specific knowledge test ³⁸⁹	SAQ/TF format, clinical vignette			Y ³⁸⁹	Y ³⁸⁹	N ³⁸⁹	-	-	-	-
Supracondylar fracture management test ³¹⁸	MCQ			-	-	Y ³¹⁸	Y ³¹⁸	-	-	-

2. OBJECTIVE ASSESSMENT OF TECHNICAL SKILL										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
A) TASK SPECIFIC CHECKLISTS										
OSATS Task Specific Checklist ^{319 322-324 351 358 367 385 387 391 392 404 405}	Checklist used to evaluate performance of individual component parts of a procedure – binary yes/no descriptors	Can be used by non-expert assessors. Useful for teaching trainees sequencing of procedural steps	Quality of performance and outcome are not captured. Rigid binary scoring does not allow for acceptable deviation from standard procedural steps. Early ceiling effect.	Y ³⁸⁵	Y ³⁸⁵	Y ^{322-324 358 367 385 387 391 392 404} N ³⁸⁹	Y ^{367 385 387 391 392}	Y ^{385 387 404 405} N ³⁹²	Y ^{322 323 385}	Y ^{324 392}
Non-OSATS Task Specific Checklist ^{294 315 329 363 375 378 400}	Procedure deconstructed into steps, often by Delphi consensus. Largely binary descriptors.	As above	As above	Y ^{294 329 375 400}	Y ^{294 329 375 400}	Y ^{329 375 400} M ³¹⁵ N ³⁶³	Y ^{329 375}	Y ^{329 378 406}	Y ^{329 400}	-
B) GLOBAL RATING SCALES										
OSATS Global Rating Scale ^{322-324 351 356 358 359 367 383 385 387 389 391 405}	Objective assessment of 7 generic open surgical skill domains (respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants). 5 point Likert scale with middle and extremes anchored by explicit descriptors ²⁹⁹	Captures quality of performance, not procedure specific, can assess complex procedures where there is more than one acceptable method. Can	Need expert surgeon evaluators. More time consuming. Does not assess skills specific to procedure	Y ³⁸⁵	Y ³⁸⁵	Y ^{296 322-324 356 358 359 367 385 387 391 392} N ^{383 389}	Y ^{367 385 387 392} N ³⁸³	Y ^{356 385 405} N ³⁹²	Y ^{322 323 385}	Y ^{324 392}

		discriminate between 'competent' and 'expert' performance.								
Non-OSATS Global Rating Scale ^{329 375}	5-point scale corresponding to Dreyfus model of skill acquisition (novice, advanced beginner, competent, proficient, expert)	Captures quality of performance. Quick.	As above	-	-	Y ^{329 375}	Y ³⁷⁵	Y ³²⁹	-	Y ³⁷⁵
Non-OSATS procedure specific GRS A ^{379 404} and B ³¹⁵	A) 2 item GRS on 5 point scale (1=very poor, 2 = clearly superior), items are overall performance and quality of final product. B) 7 global impression items evaluated on an adjectival scale from 0 to 4 (4 best)	Captures quality of performance	As above	-	Y ³¹⁵	Y ^{379 404} N ³¹⁵	Y ³¹⁵	Y ³¹⁵ 404	-	-

C) ARTHROSCOPIC OBJECTIVE SKILLS ASSESSMENT TOOLS										
Basic Arthroscopic Knee Skill Scoring System (BAKSSS) ^{334 364 365 371 379}	Objective assessment of 10 arthroscopic skill domains. 5 point Likert scale with middle descriptive anchors	Captures quality of performance. Flexible in time, can be applied live or to recorded footage. GRS component appears to be reliable when used by inexperienced raters(83)	Time consuming, does not assess skills specific to the procedure. Consent/knowledge domain may not be appropriate for technical skills assessment	Y ^{334 364 365}	Y ^{334*} 364 365 373	Y ^{334 364 371 379}	Y ^{364 365 371 379}	Y ³⁷¹ M ^{373*}	Y ³⁶⁴	-
Arthroscopic Skills Assessment Form ³⁶⁸	7 area TSC, points awarded for speed and deducted for cartilage injury. Scope score + time score = total score (max 100)	Values procedural care, timeliness and avoidance of injury. Can probably be used in-vivo	Needs expert evaluators	Y ³⁶⁸	Y ³⁶⁸	Y ³⁶⁸	-	-	-	-
Objective Assessment of Arthroscopic Skills (OOAS) ³⁷⁷	8 domain GRS with end and centre descriptive anchors	Rich descriptors, could detect subtle improvement, can show mastery	Anchor descriptors different for each domain and could be off putting for user	Y ³⁷⁷	Y ³⁷⁷	Y ³⁷⁷	-	Y ³⁷⁷	Y ³⁷⁷	-

Arthroscopic Surgical Skill Evaluation Tool (ASSET) ^{308 313 314 316 317 328 329 353 367 370 371 375 381}	Objective assessment of 8 generic ³⁷¹ arthroscopic skill domains, using 5 point Likert scale with middle anchors	Discriminates between ‘novice’, ‘competent’ and ‘expert’. Captures quality of performance. Generalisable to multiple arthroscopic procedures. Potential for use in high stakes pass-fail assessments.	Needs expert evaluators. More time consuming. Does not assess skills specific to procedure.	Y ³⁷⁵	Y ³⁷⁰ 375 381	Y ^{256 308} 313 316 317 328 329 353 367 370 371 375 381	Y ^{256 316} 328 329 367 370 371 375 381	Y ³²⁸ 329 353 370 371 381	Y ³²⁸ 329 370 381	Y ³⁷⁵
Injury Grading Index Performance Scale ²⁹⁵	Psychometric scale graded from 1-10(1 best) , similar to VAS, to subjectively evaluate for arthroscopic handling skills and intra-articular injury potential	Quick, easy to use, low cost	Requires expert assessor. Injury severity descriptors not defined, potential for subjectivity	-	-	Y ²⁹⁵	Y ²⁹⁵	-	-	-
Aggregate arthroscopic skills score ²⁹³	Score based on PT and deductions for technical errors (total score = timing score – penalty score)	May be educationally beneficial, score rewards efficiency and precision	Requires expert assessors, limited to simulated environment	Y ²⁹³	Y ²⁹³	Y ²⁹³	N	Y ²⁹³	Y ²⁹³	Y ²⁹³
D) OPEN SURGERY OBJECTIVE SKILLS ASSESSMENT TOOLS										
O-SCORE: Ottawa Surgical Competency Operating Room Evaluation ^{387 395} Ψ	11-item tool. 8 items rated on a 5-point competency scale. 1 item assessing overall procedural competence, 2 feedback items	Can be used on wide range of open and arthroscopic procedures	Needs expert assessors	Y ³⁹⁵	Y ³⁹⁵	Y ^{387 395}	Y ³⁸⁷	Y ^{387 395}	Y ³⁹⁵	Y ³⁹⁵

Generic Operative Supervised Learning Event (GOSLE) ³⁹⁴ Ψ	8-point global rating scale of construct aligned descriptors, 3 boxes for free text feedback; 2 are compulsory (reinforcing areas of good practice and areas for improvement)	Quick to complete, can be used in live or simulated setting, not procedure specific, requires minimal training in its use, has educational impact	Expert assessors	Y ³⁹⁴	Y ³⁹⁴	Y ³⁹⁴	Y ³⁹⁴	Y ³⁹⁴	Y ³⁹⁴	Y ³⁹⁴
Procedure Based Assessment ^{296 393 398}	Web or paper based assessment completed by trainer and trainee. Principle summative assessment tool in UK training	Low-cost, use in simulated or live environment, encourages trainee reflection and discussion	Cost to trainees (web subscription), need expert assessors. Blunt descriptors cannot distinguish mastery or higher order skills	-	-	Y∞ ^{296 393}	Y∞ ^{296 393}	Y∞ ^{393 398}	Y∞ ³⁹³	Y∞ ³⁹³ M ³⁹⁸

Ψ May also be used for arthroscopic procedures

* Olson study³⁷³; GRS component only

∞ Beard study³⁹³; involved multiple surgical specialities, by-speciality data was not presented

3. PROCEDURE SPECIFIC RATING SCALES										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
A) ARTHROSCOPIC PROCEDURES										
Shoulder OPAT (Objective Practical Assessment Tool) ³⁸²	Objective assessment across 4 domains; EUA, procedural standards, glenohumeral joint arthroscopy and bursoscopy. Individual step descriptors within each domain awarded points on graded scale. Global summary level 1-5	Quick to complete, addresses perceived shortcomings of PBA ³⁸² . Educational impact.	Weak inter-rater reliability ³⁸² . Needs expert assessors	Y ³⁸²	Y ³⁸²	Y ³⁸²	Y ³⁸²	M ³⁸²	Y ³⁸²	Y ³⁸²
Global Rating Scale for Shoulder Arthroscopy (GRSSA) ³⁸⁰	Objective assessment across 6 domains on a scale of 1-5 with descriptive anchors	Quick to complete, can be used in a live or remote setting	Expert assessors. Limited scope for educational impact	Y ³⁸⁰	Y ³⁸⁰	Y ³⁸⁰	-	N ³⁸⁰	-	-
Diagnostic Shoulder Arthroscopy checklist and Diagnostic Knee Arthroscopy Checklist ³⁷²	Checklist used to evaluate performance of individual component parts of a procedure – binary yes/no	Quick to complete, can be non-expert assessor	Does not evaluate quality of performance. May over-value speed	Y ³⁷²	Y ³⁷²	Y ³⁷²	-	-	-	-
Imperial Global Arthroscopy	9 domains with 5 point Likert scale with middle and end anchors, and final global assessment score	Can be used in-vivo or in simulated setting, and independent of time or place	Needs expert assessors	Y ³⁶⁶	Y ³⁶⁶	Y ³⁶⁶ ₃₇₁	Y ³⁶⁶ ₃₇₁	Y ³⁶⁶ ₃₇₁	Y ³⁶⁶	-

Rating Scale (IGARS) ^{366 371}										
Bankart Procedure Metrics score ^{307 400}	45 steps in 13 phases, Y/N for performance. Errors scored (77 possible errors, 20 sentinel)	Could be used in vivo	Needs expert assessors	Y Δ ²⁹¹	Y Δ ²⁹¹	Y ³⁰⁷	-	Y ³⁷⁸ ₄₀₀	Y ⁴⁰⁰	Y ⁴⁰⁰
B) OPEN SURGICAL PROCEDURES										
ACL error-focussed scale ³²⁷	TSC with 9 critical steps, and GRS with error focus	Could be used in vivo	Needs expert assessors. Unclear how was adapted for use	-	-	Y ³²⁷	-	-	-	-
DHS guidewire placement scoring system ⁴⁰⁷	7 domains, weighted according to the importance of the step. Maximum score 100	Could be used in vivo	Limited validity evidence, unclear how was developed and used.	-	-	Y ⁴⁰⁷ Φ	-	-	-	-
Lumbar pedicle screw placement scoring system ³⁶⁰	8 domains, weighted according to the importance of the step. Maximum score 100	Could be used in vivo	Limited validity evidence. Development unclear. Evidence limited to computer based simulation	-	-	Y ³⁶⁰ Φ	-	-	-	-
Hip fracture performance score ^{348 388}	Weighted score based on SDM, expressed as 'percentage of maximum' to account for variability between procedures	Easy, quick, automated, no requirement for assessor	Restricted to simulator	-	-	Y ³⁸⁸ N ³⁴⁸	-	Y ³⁸⁸	-	-
Supracondylar fracture	15 item TSC of key procedural steps with descriptors, binary scoring	Easy, quick could use in live theatre or simulator	Needs expert assessor	-	-	Y ³¹⁸	-	-	-	-

pinning checklist ³¹⁸											
----------------------------------	--	--	--	--	--	--	--	--	--	--	--

ΔValidity evidence from Angelo 2015 study, which was not included in the review as it includes consultants only

ΦLimited description, details of validity testing not reported fully

4. INDIVIDUAL PROCEDURAL METRICS										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
Final Product Analysis	Objective assessment of final product quality e.g. Guidewire position ^{351 362} , tip-apex distance ^{348 352} , wire-tip to subchondral distance ³²⁰ , screw position ^{255 264 326 349 361} , articular reduction error ^{363 383} , biomechanical properties ^{383 386} , articular surface congruency ³⁶³ , drill/probe accuracy ³⁵⁵ , plunge depth ³²¹ , palmar tilt ³⁵⁸	Easy to measure in simulated setting can be objective	Hard to measure in-vivo, case-specific, assessor rating can be subjective. Need expert rater	Y ³⁶¹	Y ^{351 361}	Y ^{255 320 326 348 351 361 362 386} M ^{321 352 383} N ^{349 355 358 363 374}	Y ³⁶¹ N ³⁸³	-	-	Y ^{264 351}

Fracture Fixation Assessment Tool ³⁹⁶	Objective assessment of end product using radiographs across 4 domains; reduction, stability, implant position, impression (0-3 points max for each domain) ³⁹⁶	Can be used in any time or place setting. Can be used for service evaluation as a well as training	Need expert assessor	-	Y ³⁹⁶	-	-	Y ³⁹⁶	Y ³⁹⁶	-
Error Rating ^{335 350 357}	Reports frequency of errors i.e. soft tissue/bony collisions	Can be used in any setting. Binary measure – easy to use	Need expert assessor. May be more suited to arthroscopic procedures in a simulated setting	-	Y ³³⁵	Y ^{335 350 357}	Y ³³⁵	-	-	-
Iatrogenic Index ³²⁵	Score 1 to 5 (1 best) based on intra- and extra-osseous aspect of trial and final screw trajectories	Can be used in simulated and live settings	Need expert assessor. Procedure specific	-	-	N ³²⁵	-	-	-	-
Image Intensifier use ^{320 325 350 362 388} /fluoroscopy time ^{348 351 352 388}	Number of radiographs used intra-operatively/time under fluoroscopy	Low cost, easy, can be used in simulated and live setting. Does not require expert assessor	Does not account for procedure complexity	-	-	Y ^{320 325 348 350 352 362 408} N ^{351 388}	-	-	-	-
Procedure Time ^{290 292 294 298 306 310 311 314 316 317 319 326 330-332 335-339 341 342 344-347 350-354 357 362 363 368 376 388 391 392 399}	Time for procedure or task completion	Easy to measure in simulated and in-vivo setting	Assumes speed = proficiency. In vivo patient and staff factors can affect procedure time	-	-	Y ^{290 292 294 298 306 310 311 326 332 335-339 341 342 344-346 351 353 354 357 368 376 399} M ^{314 330 331 392} N ³¹⁶	Y ^{306 316 338 376} N ³⁹¹	-	-	-

						317 319 347 350 352 362 363 388 391				
Cumulative Summation Test for Learning Curve (LC-CUSUM) ³⁷⁹	Derived from cumulative scores to determine when a trainee has reached a pre-defined level of competence	Allows for quantitative individualised assessment of learning and a graphical representation of the learning curve. Data can come from live or simulated setting	Requires sequential data points in series. Needs specialist statistician input to implement. Consensus required on pass threshold	-	-	-	-	-	Y ³⁷⁹	-

5. MOVEMENT ANALYSIS										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
Hand Motion Analysis Systems, hand and instrument motion analysis ^{312 333 341 363-365 376}	Sensors attached to the dorsum of surgeons hands/instrument tips record speed, distance and number of movements, can generate efficiency and economy scores	Sophisticated data profile enabling detection of subtle improvement in performance. May be able to measure attainment of mastery	Difficult to use in-vivo because of sterility considerations. Invasive for the surgeon being assessed	-	-	Y ^{312 333 341 363-365 376} M ³⁵⁶	Y ^{333 356 364 365 376}	Y ³⁶⁵	Y ³⁶⁵	-

Elbow Motion Analysis ²⁹⁸	Elbow-worn sensors generate data on motion metrics; number of hand movements, number of minor movements, smoothness and time	As above. More feasible to use in-vivo as sensors sit at elbows (above sterile area)	Can be used in vivo	-	-	Y ²⁹⁸	-	-	Y ²⁹⁸	-
Gaze tracking using eye movements ³⁰⁶	Glasses worn by surgeon with sensors to track eye movements. Expressed as time and proportion fixed on screen vs. hands	Could be used in vivo. Sophisticated data profile enabling subtle performance improvement	Invasive for surgeon being assessed	-	-	Y ³⁰⁶	Y ³⁰⁶	-	Y ³⁰⁶	-
Hand position checking ^{292 364}	Number of occurrences of hand position checking during procedure is recorded (frequency is inversely proportional to experience)	Low/no-cost, easy, non-expert assessor, can be used in simulation and in-vivo	Does not assess quality of performance or outcome	-	-	Y ^{292 364}	Y ^{292 364}	-	Y ²⁹²	-
Instrument loss ³⁶⁴	Number of instances during which the tip of the arthroscopy probe was not visible on the arthroscopy display unit. Assesses visuospatial awareness and fine motor dexterity	Free of cost, easy to measure, can be used in simulation and in-vivo	Does not assess quality of performance or outcome	-	-	Y ³⁶⁴	Y ³⁶⁴	-	-	-
Triangulation time ³⁶⁴	Mean duration of instrument loss episodes during procedure (total duration of instrument loss in seconds divided by the number of instances of instrument loss)	Free of cost	Requires precise timekeeping, not easy for scrubbed trainer. Easier if video recorded procedure	-	-	Y ³⁶⁴	Y ³⁶⁴	-	-	-
Instrument path length/ratio ^{311 314 332 339 340 342 345 346 353 399}	Composite measure of total distance travelled by camera and probe during an arthroscopic procedure, or ratio of measured path relative to the ideal path. Gives a measure of movement economy	Easy to obtain	Cannot be measured in vivo	-	-	Y ^{311 332 339 340 342 345 346 353 399} M ³¹⁴	Y ^{332 342 345}	-	-	-
Instrument collision force ^{330 345}	Precision, distribution of forces applied to the joint surface and	Sophisticated data profile, can see	Cannot be measured in vivo	Y ³³⁰	-	Y ³⁴⁵ M ³³⁰	-	-	-	Y ³³⁰

	efficiency of task completion can be characterised.	subtle improvements in performance								
Collisions and injuries ^{311 332 338-340 344 353}	Number of times the probe/arthroscope contacts tissue, and injuries defined as contacts beyond a threshold force	Easy to obtain	Cannot be measured in vivo	-	-	Y ^{311 332 338 339} M ³⁵³ N ³⁴⁰	-	-	-	-

6. PSYCHOMOTOR TESTING										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
Grooved Pegboard Test (Lafayette Instrument, Lafayette, IN) ^{309 390 397}	Time to complete the 25-hole board is recorded for dominant and non-dominant hand. Originally developed to test the dexterity of assembly line workers	Cheap, easy, portable, non-expert assessor	Results do not relate to patient or operative outcome	-	-	Y ³⁹⁰ M ³⁰⁹	Y ³⁹⁰ N ³⁹⁷	-	Y ³⁹⁰	-
Arthroscopic Knot Trainer ^{310 347}	Non-anatomical, low-fidelity bench model with spring loaded eyelets to simulate tissue under tension during surgical repair	As above	Cannot use in-vivo	Y ³⁴⁷	Y ³⁴⁷	Y ³¹⁰ 347	Y ³¹⁰	-	-	Y ³⁴⁷

The Crawford Small Parts Dexterity Test ³⁹⁷	Tests hand eye co-ordination. Has application in the clinical setting for hand rehab by OT's. Part 1 measures dexterity in using forceps to insert metal pins in close-fitting holes and place collars over the pins. Part 2 measures dexterity in starting and setting screws in threaded holes in a metal plate	As above	As above	-	-	-	M ³⁹⁷	-	-	-
Steadiness Hole Test ³⁹⁷	Measures hand steadiness as the ability to hold a metal stylus in a fixed position within holes of different diameters without touching the sides	As above	As above	-	-	-	M ³⁹⁷	-	-	-
Bennett hand-tool test ³⁹⁷	Measures dexterity in handling ordinary mechanical tools. Developed for application in industrial apprentice training. The test consists of tools and two uprights with bolts, the objective is to disassemble the bolts from one upright and re-assemble them on the corresponding rows on the other test-board uprights	As above	As above	-	-	-	N ³⁹⁷	-	-	-
Minnesota rate manipulation test ³⁹⁷	Tests hand-eye co-ordination, originally developed as pre-employment screening tool, now mainly clinical application in rehab. Uses a hole-punched board and blocks, 5 sub-scores for complete score	As above	As above	-	-	-	N	-	-	-
Autoscoring Mirror Tracer(Lafayette Instrument, Lafayette, IN) ³⁰⁹	Tests reversal ability. Participants trace a star pattern whilst watching only a mirror image	As above	As above	-	-	M ³⁰⁹	-	-	-	-

Purdue Pegboard Test (Lafayette Instrument, Lafayette, IN) ³⁰⁹	Measures gross hand and finger movements in assembly tasks	As above	As above	-	-	Y ³⁰⁹	-	-	-	-
O'Connor Tweezer Dexterity Test (Lafayette Instrument, Lafayette, IN) ³⁰⁹	Tests fine manual dexterity, participants have to use tweezers to place 1/16" pins in holes	As above	As above	-	-	Y ³⁰⁹	-	-	-	-
Etch-a-Sketch with overlay (Lafayette Instrument, Lafayette, IN) ³⁰⁹	Tests two hand and hand-eye coordination, participants have to trace a pattern on a standard Etch-a-Sketch	As above	As above	-	-	M ³⁰⁹	-	-	-	-
Two-arm coordination test (Lafayette Instrument, Lafayette, IN) ³⁰⁹	Tests coordination and balance. Participants trace an anodized star pattern with a two-handed triangular pointer.	As above	As above	-	-	M ³⁰⁹	-	-	-	-

7. SUBJECTIVE ASSESSMENTS										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
Pass/Fail judgement ^{385 391 392}	Subjective global judgement by expert assessor as to whether trainee completed the tasks and performed each part in a safe and controlled manner	Any location and procedure	Expert assessors only. Subjective. Not reliable.	-	-	Y ^{385 391 392}	Y ^{385 392}	N ³⁸⁵	-	-
Video Feedback ³⁸⁴	Operative performance is video recorded and then reviewed by the trainee, either with self –reflection or expert feedback	Encourages reflection, learning from own mistakes, used effectively in other industries (e.g. athletics)	Requires camera equipment set up, ethical issues with filming real operations	-	-	N ³⁸⁴	N ³⁸⁴	-	Y ³⁸⁴	N ³⁸⁴

Y = Evidence demonstrated for utility property

M = Mixed evidence demonstrated (positive and negative findings for utility property within same study)

N = No evidence demonstrated for utility property

Abbreviations

EUA = Examination under anaesthetic

OT = Occupational Therapist

VAS = Visual Analogue Scale

5.5 Discussion

Robust assessment of competency and operative skill in T&O is a topical issue in training for the reasons outlined in the introduction. The primary goals of surgical competency assessment are to provide a platform for learning through feedback, to make summative judgements about capability and progression through training, to maintain standards within the profession and ultimately to protect patients from incompetent surgeons²⁹⁹.

This review is, at the time of writing, the first systematic analysis of the tools currently available for assessing technical skill and operative competency in T&O surgical training.

The results show that none of the tools currently used for assessing technical skill in T&O surgical training fulfil the Norcini criteria for effective assessment⁴⁰⁹ (validity; reproducibility; equivalence; feasibility; educational effect; catalytic effect and acceptability). There is a similar deficiency of utility evidence in general surgery⁴¹⁰ and vascular surgery²⁹⁹, which like T&O are facing the same challenges in moving towards a competency based approach to training²⁹⁹ and needing robust methods of assessing technical skill and competency.

The results of the review show that checklists and global rating scales are the most commonly used tools for technical skills assessment in T&O (table 10, section 2a-b). Checklists deconstruct a task into discrete steps, and may in and of themselves have educational value for teaching procedural sequencing to novice surgeons-in-training. They do not capture quality of performance, and the rigid binary scoring does not allow for deviation from the prescribed step-by-step process in procedures where there may be more than one accepted method. Checklists also suffer from the problem of having an early ceiling effect²⁹⁹. Checklists can be administered by non-expert assessors, and judgement on performance can be made either live *in situ* or retrospectively from video footage. Checklists can also be easily used in both the simulated and live theatre environment. They show reasonable construct validity³⁵⁸^{367 385 387}, concurrent validity^{329 367 385 391 392} and reliability^{329 378 400}. With these

limitations in mind, checklists are perhaps most appropriate for novice learners in a formative setting²⁹⁹.

Global rating scales for surgical assessment use generic domains with a Likert-type scale and descriptive anchors to capture the quality of performance^{351 356 383}. They are generalisable across different procedures, and can be used to assess complex procedures and those where there is more than one accepted technique. Global rating scales can discriminate between ‘competent’ and ‘expert’ performance, and the results show there are many studies demonstrating evidence of content validity^{315 385}, concurrent validity^{315 367 375 385 387 392}, and reliability^{315 329 356 385}. Global rating scales have the disadvantage of requiring expert surgeon evaluators to be administered effectively and are more time-consuming to use for assessment. They may also be susceptible to a degree of assessor bias, as the global domains of assessment such as ‘instrument handling’ and ‘respect for tissue’ are inherently quite subjective, although the inter-observer reliability was shown to be generally high^{294 356 385}. The ability of global rating scales to distinguish between all levels of performance and the absence of any ceiling effect makes it useful for high stakes, summative assessment²⁹⁹, and assessment of advanced residents nearing the end of training.

There have been several novel objective assessment tools that have been developed to combine task specific checklists with a global rating scale. They combine the known benefits of both types of assessment into a single tool. The review results suggest that the most promising front-runners amongst these are the Arthroscopic Surgical Skill Evaluation Tool (ASSET)^{328 329 367}, which combines a task specific checklist with an eight-domain global rating scale with end and middle descriptive anchors (table 10 section 2c), and the Objective Structured Assessment of Technical Skills (OSATS)^{319 383} (table 10 section 2b). The ASSET is by definition restricted to assessing arthroscopic procedures, but both tools have an impressive and growing body of evidence across all domains of the utility index (table 10). This novel hybrid approach does however have the disadvantage that they can become quite long and burdensome to complete, which might negatively impact their feasibility and acceptability in a busy workplace where assessment of training conflicts with service pressures.

The OSATS is currently in use in UK training programmes in Obstetrics and Gynaecology⁴¹¹ and Ophthalmology⁴¹² and has proven popular with trainees⁴¹³. OSATS can capture the ‘quality’ of performance and can distinguish competence from mastery and the stages of progression between. There were several studies in this review which demonstrated the validity, reliability, feasibility and educational value of the OSATS in T&O in the simulated setting (table 10, columns 5-11). Further work is required to assess its utility in the live operating theatre before it can be recommended for use in training.

The results also show that there are a variety of unique procedure specific rating scales that have been developed for both open^{318 327 348 360 388 407} and arthroscopic^{307 366 371 372 380 382} T&O procedures (table 10 section 3a-b). Most of these are in the early stages of validation and are largely restricted to use in the research setting. Procedure specific checklists are not particularly practical for use in the live operating theatre environment given the variety of procedures that are undertaken within a typical rotation or placement, a single generic tool that may be applied to the assessment of all relevant training procedures is more appealing from a feasibility and acceptability perspective.

Motion analysis (table 10 section 5) is also another promising area for assessing technical skill. Motion analysis appears to be of particular value in assessing performance in arthroscopy, and several studies in this review demonstrated its utility^{292 298 306 312 333 341 356 364 365 376}. Its use to date has been largely restricted to the research setting, and so further work on transfer validity and potential educational impact is required. Some of the obvious barriers to use in the live operating theatre environment, such as sterility concerns with surgeons wearing motion tracking sensors, have been creatively mitigated by using elbow instead of hand-mounted sensors²⁹⁸. Hand motion analysis techniques can generate a sophisticated data profile that can detect subtle improvement in surgical performance, and may therefore be able to measure the attainment of mastery with a high degree of resolution. Other intra-operative motion parameters such as gaze tracking³⁰⁶, triangulation time³⁶⁴, instrument path length^{311 314 332 339 340 342 345 346 353 399} and collisions^{330 345} have demonstrated construct validity and feasibility in the simulated environment (table 10, section 5, columns 7 and 10), but these are unlikely to be useful in the live

operating theatre as most of these measurements are derived from the simulator itself.

The utility index appraisal (table 10) shows that individual procedural metrics can also be used to effectively assess technical skill (table 10 section 4). Final product analysis provides an objective assessment of final procedural quality, from which technical proficiency can be inferred. Examples of final product analysis assessment include Dynamic Hip Screw tip-apex distance^{348 352}, screw position^{255 264 326 349 361}, and articular congruency in complex articular fracture surgery^{363 383}.

Orthopaedics has an advantage over other surgical specialties in the use of final product analysis due to the routine use of intra-/ post-operative radiographs from which relevant real-life final product analysis metrics such as implant position can easily be measured. Final product analysis is also objective and quite easy and efficient to obtain. A non-specialist assessor (who has been appropriately trained) can make the measurements at a time and setting remote from the procedure. In the simulated/research setting, invasive final product analysis measures such as biomechanical testing of a fracture construct can be used to assess procedural success. Final product analysis is appealing as an assessment method as it relates technical performance to real world, clinically relevant measures of operative success. It may be particularly appropriate in T&O procedures where there is a clear evidence base of implant position predicting clinical outcome⁴¹⁴. The construct validity picture is however rather mixed, with almost as many studies refuting its construct validity^{349 355 358 363 374} as those demonstrating it^{255 320 326 348 351 361 362 386} and no evidence as yet of reliability or educational impact.

Procedure time has been extensively used as a metric to assess technical skill in the studies included in the review. It has the advantage of being quick, easy and cheap to measure in both the simulated and *in vivo* setting. Procedure times (as knife-to-skin to wound-closure, or equivalent) are routinely measured in operating theatres in T&O as part of theatre efficiency measurement auditing by hospital management. This data on real-world performance is therefore already available without any additional resource requirement.

Using procedure time to measure technical skill does however rely on the intuitive assumption that speed equates to proficiency. This is potentially problematic, as external patient and staff factors beyond the surgeon's control could easily influence procedure time in the operating theatre. Procedure time also gives no indication of quality of performance; a procedure may be fast because the surgeon was a masterfully efficient operator, or alternatively they could just as easily have rushed the procedure or been careless. The construct and concurrent validity evidence for procedure time is mixed, with many studies showing it can discriminate between experience level^{290 292 294 298 306 310 311 326 332 335-339 341 342 344-346 351 353 354 357 368 376 399}, and performs well against other types of assessment^{306 316 338 376}, and several other studies showing it cannot^{316 317 319 347 350 352 362 363 388 391}. For these reasons, final product analysis and procedure times are therefore unlikely to be useful measures in isolation, but rather could be used as adjunctive assessments of technical proficiency.

5.5.1 Context with current assessment practice in T&O in the UK

There is growing dissatisfaction with the current technical skills assessment tools within the surgical education community^{394 415}, and an increasingly urgent need to develop evidence-based assessment methods for use within a competency-based training environment. Such tool(s) need to be generalisable to the broad range of technical and non-technical skills in T&O, should satisfy Norcini's criteria and the various domains of the utility index (described in table 10).

The procedure based assessment is the current gold-standard assessment tool used for high-stakes summative assessment in the UK training system in T&O. It is lengthy to complete, comprising 40-50 tick boxes and 12 free-text spaces³⁹⁴. The procedure based assessment was initially implemented prior to any formal validation, beyond an initial consensus setting (Delphi) process to define the domains^{394 416}.

Several years after its introduction, a large pan-surgical speciality validation study was undertaken³⁹⁸ with a particular focus on demonstrating the reliability of the rating scales³⁹⁴. Within this study, T&O is surprisingly under-represented, despite

being the second largest surgical speciality, with the totality of the procedure based assessment validity evidence in T&O being drawn from 2 orthopaedic procedures involving just 7 trainees. Subsequent validation work followed on from this study using more traditional frameworks in general and vascular surgery. This has shown the procedure based assessment is both a valid and reliable measure of performance³⁹⁴ and responsive to change³⁹⁴, but there remains a deficiency of evidence for its utility in orthopaedics (table 10 section 2d).

In addition to the lack of validation evidence in T&O, the engagement with the procedure based assessment in the T&O trainee population has been poor³⁹⁴, and it remains unpopular⁴¹⁵. A national survey of T&O trainee attitudes towards procedure based assessments found that more than half agreed or strongly agreed with the statement ‘completing PBA’s is nothing but a form filling exercise’, 60% agreed or strongly agreed that there are ‘barriers to the successful use of PBA’s by trainees’⁴¹⁵ and just a third believed they should be used for high stakes assessment in training, such as the Annual Review of Competence Progression⁴¹⁵. Further work has explored the reasons for the poor engagement and enduring unpopularity of the procedure based assessment. This has shown that it is perceived to be burdensome to complete in a climate of trainee assessment fatigue, and with a coarse rating scale of blunt, binary descriptors it cannot distinguish mastery⁴¹⁵.

The procedure based assessment was among the earliest formal tools for technical skill assessment in T&O surgical training. As such its creators deserve recognition for beginning the process of objectively assessing technical skill in surgical training. I believe that the results of this review show that the procedure based assessment is no longer fit for purpose and should not be used for summative skills assessment in a modern competency based assessment training environment.

The OSATS and the ASSET appear to show promise as potential replacements to the procedure based assessment as gold-standard assessment tools in T&O training. Efforts toward validation work on these tools, with a particular focus on their use in the live operating theatre, should be continued.

5.5.2 Strengths and Limitations

An advantage of this review is it is the first comprehensive evaluation of the current tools for assessing technical skill acquisition and operative competence in T&O.

This review is limited to the assessment of technical skills in T&O, and I deliberately did not consider the assessment of non-technical skills for surgeons (NOTSS). Non-technical skills are undoubtedly an essential dimension of surgical competence, as per the famous surgical aphorism ‘you can teach a monkey to operate’²⁹⁹, and it is rightly beginning to receive attention in the surgical education literature⁴¹⁷. The perfect technical skills assessment tool is therefore never going to be usable in isolation to comprehensively assess competence, but rather should form a key part of a battery of evidence-based assessment tools. A similar review is needed for the current use of NOTSS in T&O, and this is an area for further work.

5.6 Conclusion

The utility evidence for the procedure based assessment in T&O is inadequate to support its continued use in summative high stakes assessment of competency. An assessment tool that is generalisable to the broad range of technical and non-technical skills relevant to T&O that satisfies the utility criteria, is cost-effective and feasible requires development.

5.7 Reflections

With the benefit of the experience gained in conducting the first systematic review in this thesis (Chapter 3), I found the execution of this one considerably easier. I used software to organize the review process, and found this to be very useful. That said, extracting the data was an enormous amount of work (as evidenced by the size of the tables) and took me several months to do.

I was pleased that the review was accepted for publication by the American Journal of Bone and Joint Surgery 'Reviews', with minimal revisions. One reviewer described it as an 'important benchmark' paper in the field of surgical education, which I was absolutely delighted by.

Having done this work I now have a clear view that final product analysis of implant position on intra/post-operative radiographs is a suitable and sensible primary outcome measure to assess the real-world impact of cadaveric simulation on surgical trainees' performance. The next challenge is working out what to measure on the radiographs, and what the acceptable cut-off values should be. An exploration of this will form the next chapter of this thesis.

6. Development of an objective outcome measure of real-world technical operative success using post-operative radiographs

In this chapter I will undertake a Delphi exercise to try and achieve expert consensus on the radiographic features of technical success for Dynamic Hip Screw, hip hemiarthroplasty, and ankle fracture fixation. With the exception of tip-apex distance in DHS, there are currently no agreed criteria for what the optimal implant position is for these procedures. Establishing what constitutes an optimally-placed implant will enable me to develop a set of sensible outcome measures for the trial.

Declarations

I utilized, with permission, the Ankle Injury Rehabilitation (AIR) Trial (ISRCTN17809322) radiology manual as a starting point for developing radiographic outcome measures for ankle fracture fixation

Mrs Yessica Diez-Davies enhanced my diagrams using professional graphic design software

I sought advice from Professor Richard King to provide clinical context to the findings of my scoping literature search in determining what might be important radiographic outcome measures after hip fracture surgery.

The results of this chapter reflect the opinions of the Delphi exercise participants, who are all consultant orthopaedic surgeons at University Hospital Coventry & Warwickshire, which is a large regional trauma centre.

6.1 Introduction

6.1.1 Rationale

To enable assessment of the impact of CST on patient outcomes (Kirkpatrick level 4), a suitable, objective, patient-related outcome measure is required. In clinical trials, Patient Reported Outcome Measures (PROM's) are generally considered as the gold standard, patient-centred outcome measure^{418 419}. In educational research, it is probably neither appropriate nor practical to rely on PROMs as the primary outcome measure. This is because there are numerous potential confounding variables from the time of operation to the time of PROM assessment that could influence the outcome other than the operating surgeon's skill level. These include hospital-related factors⁴²⁰ (such as post-operative nursing care and physiotherapy rehabilitation), and patient-related factors^{421 422} (such as psychological state, socioeconomic status, expectations of surgery and functional demand). These factors are therefore likely to obscure surgical skill-related factors in assessing outcome. A diagrammatic explanation of this problem is shown in Figure 21 below.

Another way of measuring the real world impact of the training intervention would be to assess clinical outcome by way of patient morbidity and mortality. This suffers the same set of problems as PROMs, in the sense that there are myriad factors that influence morbidity and mortality other than surgical performance, and controlling for these would be challenging, and the chances of seeing an effect of the training is probably quite small (figure 21).

Final Product Analysis (FPA) refers to the technique of scoring performance based on the quality of the final product. The product in this instance is the positioning of the implant, as measured from the post-operative radiograph.

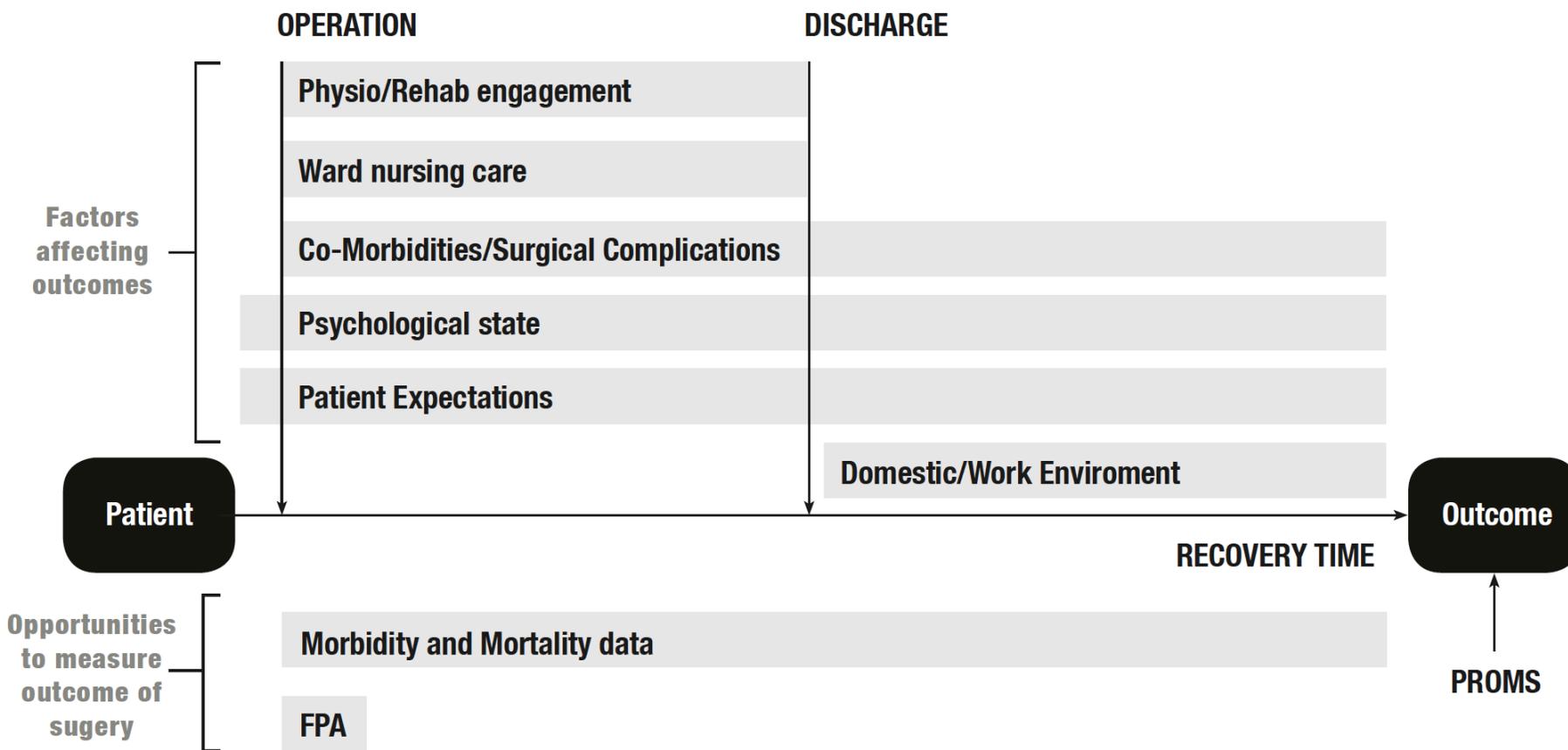


Figure 21. Schematic overview of patient post-operative recovery timeline and the factors that can influence outcome from surgery

FPA = Final Product Analysis

PROMS = Patient Reported Outcome Measures

6.1.2 Final Product Analysis as a measure of technical operative success

All three of the procedures under study in the trial routinely have radiographs taken either intra-operatively (as is the case for DHS) or post-operatively (as is the case for hip hemiarthroplasty and ankle ORIF). I think it is reasonable to use these radiographs to undertake ‘final product analysis’ (FPA) of the procedures, as performing an operation is a complex technical skill. Hence an objective assessment of the technical outcome of surgery seems appropriate for measuring the impact of a training intervention on skill. FPA assessment is made at a time point before other factors begin to have influence on outcome, so feels like a better choice of primary outcome measure than PROM’s or morbidity/mortality data in the setting of an educational trial.

It is controversial, however, to assume that a technically successful operation as measured by an radiograph means that the patient (or their surgeon) is happy with the outcome. A detailed exploration of the relationship between objective technical success as measured by FPA and patient or surgeon satisfaction is important, but is not within the scope of this thesis. I also do not seek to develop an entire new outcome measure that meets the utility criteria discussed in chapter 5, as this would be a huge undertaking and a doctoral project in its own right. I am undertaking this Delphi exercise for the purposes of developing a sensible set of radiographic outcome measures to assess the technical skill of surgeons participating in the trial.

The ideal radiographic parameters for this purpose are those that are measurable on a post-operative radiograph, that are clinically relevant, and which are responsive to change in technical skill level (figure 22).

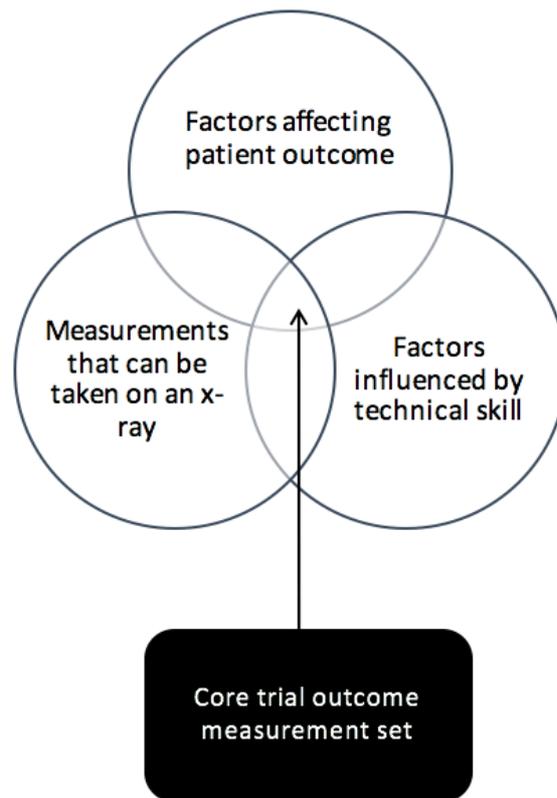


Figure 22. Venn diagram showing the required qualities of the radiological measurement parameters

6.2 Aim

The aim of this chapter is to define the outcome measures for the educational trial (chapter 7). This will be a core outcome set of measurements that can a) be taken from post-operative radiographs, b) which are influenced by technical skill, and c) which affect patient outcome (figure 22).

6.3 Consensus setting exercise

6.3.1 The Delphi process

The Delphi technique, originally developed in the 1940's, is widely used to 'systematically combine expert opinion to arrive at an informed group consensus on a complex problem'⁴²³. This is typically achieved using 'iterative rounds of sequential surveys interspersed with controlled feedback reports and the interpretation of expert opinion'⁴²³. Individuals are invited to reconsider their initial position in light of group trends, and the field of opinion is narrowed until consensus is reached. The Delphi technique has several advantages; anonymity encourages participants to interact 'free from the constraints of personality conflicts or status relations'⁴²³ and it can be conducted without participants physically meeting so is free from logistical and geographic difficulties⁴²⁴.

The technique has been criticized for not enabling 'cross-table' discussion as the participants do not physically meet⁴²⁵, although some modified Delphi approaches have been specifically developed to include a face-to-face element⁴²⁶. The e-Delphi technique (using internet based survey platforms) offers convenience, time and cost savings, and data management advantages⁴²³ as compared to traditional paper-based methods. I have used a simple e-Delphi technique in this chapter.

6.3.2 Methods

Sampling

The entire cohort of Trauma & Orthopaedic consultants in a regional Major Trauma Centre (University Hospitals Coventry & Warwickshire) were invited to take part in the Delphi survey (n=39). I included all T&O consultants in the invitation to take part, irrespective of sub specialism or involvement with on-call trauma responsibilities, as the operations of interest are all 'core' procedures in which

independent competence is required before the award of the Certificate of Completion of Training (CCT).

I therefore thought it was reasonable to include all consultants (who are by definition CCT holders) in the survey, as these consultants would all be capable of performing these procedures, and supervising/ assessing a junior colleague doing so.

Survey Design

I initially undertook a scoping review of the literature to identify the current evidence for assessing technical skill using radiographs. I found no evidence for this in the literature, and so I moved the focus of the scoping review to looking for evidence of radiographic features of the three operations of interest that predict outcome after surgery.

The radiographic features that predict a good clinical outcome and those that suggest technical success of the operation may or may not be the same (intuitively they are unlikely to be mutually exclusive), however given the lack of evidence of the latter it seemed a sensible place to start.

In designing the survey I contacted internationally recognized experts in the field of hip trauma (Martyn Parker, personal communication) and ankle fracture (Ankle Injury Management trial team, in person) to check the survey items I had developed from the literature review were in line with current leading opinion.

The survey was built using an online survey platform (Survey Monkey Inc., San Mateo, California, USA). I kept the survey deliberately short as my sample population are busy and survey-fatigued NHS professionals. Reminders were sent at 2 and 4 week intervals after initial invitation. Non-responders after the second reminder were presumed as withdrawn.

The threshold for consensus was set at 75% agreement. This was considered the cut-off for consensus, as this was the median consensus threshold in a large systematic review of Delphi methodology⁴²⁷

Round 1

There was a survey question page per operation, and the proposed measurement items were listed with binary yes/no answer options, with free text comment space.

Example question:

'Please indicate if you consider cement mantle thickness to be important for assessing the technical success of a hip hemiarthroplasty operation. Please assume you are looking at an adequate standard antero-posterior post-operative radiograph (answer options: yes/no)'

Round 2

Items that had achieved consensus in round 1 were re-presented with proposed cut-off thresholds, which were derived from the best available literature evidence. Participants were asked about the reasonableness of the proposed cut-off definitions, in a binary yes/no format.

Example cut-off question:

'90% of respondents in round 1 agreed that leg length discrepancy (LLD) in hip hemiarthroplasty was relevant. Do you think a cut-off for acceptable LLD post-op of <10mm is reasonable? (Answer options: yes/no)'

Items that had not achieved consensus were re-presented with the results of round 1 (in terms of % agreement) with additional details of supporting literature evidence.

Example re-presentation question:

'Opinion was split in round 1 as to whether lag screw position in the femoral head is important (55% in favour). The literature evidence suggests that a high anterior

screw position in the head seems to predict cut-out independently of tip-apex distance. Do you think it is worth measuring screw position in view of this? (Answer options: yes/no)'

Round 3

Items that had still not achieved the cut-off threshold for consensus were re-presented, with new threshold proposals in line with panel opinion from round 2, along with relevant supporting literature evidence to support the decisions made, where appropriate.

Example question:

'67% of participants in round 2 thought 10mm was a reasonable cut-off for acceptable leg length discrepancy after hip hemiarthroplasty. 75% panel agreement is needed for consensus. We therefore propose revising the definition of acceptable LLD to <15mm in view of panel feedback. Is this reasonable? (Answer options: yes/no)'

Items that reached consensus for inclusion in round 2 were re-presented with cut-off thresholds with supporting evidence.

The steps of the consensus setting process is shown in Figure 23.

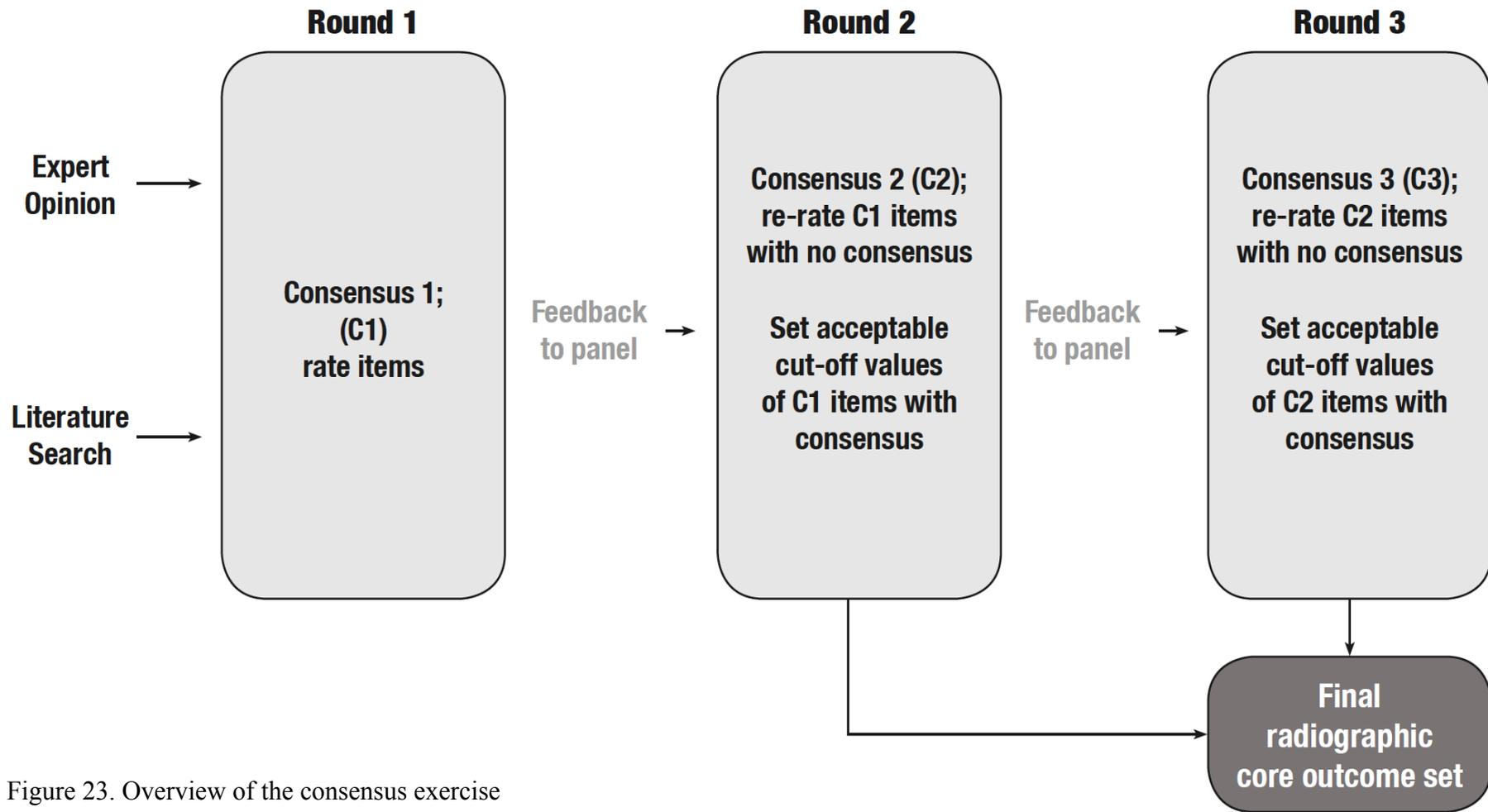


Figure 23. Overview of the consensus exercise

6.3.3 Results

Twenty-six responses to the first survey round were received. Twenty-one responses were received from round 2, and nineteen from round 3. This represents a 49% completed participation rate from the eligible cohort.

The final consensus items are shown in table 11 below. For definitions, illustrations and supporting literature evidence for these items, please see the CAD:TRAUMA Radiology Guide in appendix 11.5.

Table 11. Final items

i. Ankle Fracture Fixation

1. Included Items		
Item	Round in which consensus was achieved	% Agreement
1.1.a. Medial Clear Space (MCS)	1	88
1.1.b. Acceptable MCS = ≤ 4 mm	2	76
1.2.a. Medial Malleolar Displacement (MMD)	1	96
1.2.b. Acceptable MMD = ≤ 2 mm	3	95
1.3.a. Lateral Malleolar Displacement (LMD)	1	96
1.3.b. Acceptable LMD = ≤ 2 mm	3	95
1.4.a. Tibiofibular Clear Space (TFCS)	1	75
1.4.b. Acceptable TFCS = < 5 mm	3	100
1.5.a. Talo-Crural Angle (TCA)	3	95
1.5.b. TCA $80^{\circ} \pm 5$	3	95
1.6.a. Talar Subluxation (TS)	1	92
1.6.b. TS presence/absence	1	92

2. Excluded Items	
Item	Exclusion Reason
Talar Tilt Angle	Failed to reach consensus after 3 rounds. Inferior to Talo-Crural Angle.

ii. Dynamic Hip Screw

1. Included Items		
Item	Round in which consensus was achieved	% Agreement
2.1.a. Tip-Apex Distance (TAD)	1	88
2.1.b. Acceptable TAD < 25mm	2	90
2.2.a. Trabecular Angle (TA)	3	84
2.2.b. Acceptable TA = 150 ° -170 °	3	84
2.3.a. Lag screw position in femoral head	2	90
2.3.b. Described according to Cleveland's 9 zones; Anteroposterior : Superior/Central/Inferior. Lateral: Posterior/ Central/ Anterior	2	90
2.4.a. Plate position	1	83
2.4.b. Acceptable = plate flush with cortex on AP, no gaps seen	1	83
2.5.a. Cortical Screw Position	1	88
2.5.b. Acceptable = 8 cortex hold	1	88

2. Excluded Items	
Item	Exclusion Reason
Screws perpendicular with plate	Rejected, 68% against in Round 1

iii. Hip Hemiarthroplasty

1. Included Items		
Item	Round in which consensus was achieved	% Agreement
Leg Length Discrepancy (LLD)	1	92
Acceptable LLD = ≤ 15 mm	3	89
Femoral Stem Alignment (FSA)	1	88
Acceptable alignment = \leq or $\geq 5^\circ$ from neutral	3	95
Cement Mantle Grade	2	81
Acceptable = give grade according to Barrack's criteria	2	81
Femoral Offset	2	81
Acceptable = should be equal to native side	2	81

1. Excluded Items	
Item	Exclusion Reason
Cement thickness	Failed to reach consensus after 3 rounds. Inferior to cement mantle grade

6.4 Discussion

In attempting to define a core radiographic outcome set to measure technical skill, I had expected to be able to find existing work in the literature reporting on what features of a post-operative radiograph predicts clinical outcome. I had this expectation because in the clinical environment, post-operative radiographs are a

well-established part of routine patient care to assess the success of orthopaedic surgery. I therefore assumed that these judgements must be informed by evidence.

In the hospital morning trauma meeting, or the outpatient clinic, a T&O doctor will look at the post-operative radiograph to help decide if the operation has been successful or not. In the relative absence of proper evidence in the literature, this assessment of the radiograph appears to be a global, qualitative, ‘expert eye’ judgement. An experienced clinician can quite quickly look at an radiograph and make a judgement about the technical success of the operation. This seems to be a skill that is acquired through years of experience. In this chapter, I am attempting to distil the ‘expert eye’ into a list of criteria that can be used to judge the technical success of an operation, in an objective and reproducible manner.

With the notable exception of DHS, where TAD is widely accepted as a predictor of implant failure⁴¹⁴, I was surprised to see that there was very little evidence for radiographic factors predicting outcome for hemiarthroplasty and ankle fracture fixation. The latter was especially surprising given the relative abundance of clinical trials assessing the impact of various surgical and non-operative interventions^{428 429}.

For hemiarthroplasty, I looked further into the total hip arthroplasty (THA) literature, as a cemented hemiarthroplasty is quite literally half a THA, and hence any evidence pertinent to the outcomes of the femoral component position in THA should, intuitively, be able to be extrapolated to hemiarthroplasty. This assumption is of course complicated by patient group differences; logically the clinical outcome of a fit active 50-something year old undergoing an elective THA would be influenced to a greater extent by optimal implant position, when compared to a frail 90 year old undergoing a hemiarthroplasty for fractured neck of femur whose functional demands are low and who really ‘just’ needs pain relief and joint stability to transfer from bed-to-chair in a nursing home.

I found some good quality evidence that showed there were negative clinical implications to leg length discrepancy after THA⁴³⁰, and varus malposition of the femoral component^{431 432} and some evidence that poor cementation is correlated with early failure⁴³³ – all of which can be measured on the post-operative radiograph.

There was a similar paucity of evidence for post-operative radiographic assessment of ankle fracture fixation. Most of the evidence, as would logically be expected, relates to the precision of the reduction post-surgery, as measured by medial⁴³⁴ and tibiofibular⁴³⁵ clear spaces, malleolar reduction accuracy⁴³⁶, and talo-crural angle⁴³⁴⁴³⁷. The most recent of the ankle fracture evidence was nearly 30 years old, which suggests to me that this is a neglected research area.

Given the general lack of modern evidence on the radiographic features of technically successful fracture fixation and, anecdotally, lack of agreement amongst trauma surgeons on the same, meant that a consensus exercise seemed appropriate.

6.4.1 Clinical importance vs technical importance

All the available literature evidence, much of which I have described above, relates to radiological factors that predict clinical outcome. An additional layer of complexity in defining an outcome set lies in the fact that not everything that may be measurable on an radiograph might be clinically relevant, and not everything measurable on an radiograph might be influenced by the technical skill of the surgeon. The challenge was to find the area of overlap between these sets of variables (figure 22).

It was apparent from the free-text comments section after the first round of the survey that many participants were focused exclusively on measurable factors that influence clinical outcome, as many comments were along the lines of ‘this doesn’t really matter clinically – why bother measuring it’. This is understandable coming from a group of clinicians, who will naturally be looking at this through a very clinical-outcomes focused lens. When giving feedback from Round 1 to participants in Round 2, I reiterated that the primary objective of the exercise was to find measurements that were both clinically relevant and allowed assessment of technical skill in an educational research trial.

An additional challenge was in finding radiographic outcome measures that had an appropriately high resolution to detect small incremental gains in skill. For example, it is obviously clinically relevant as to whether a hip hemiarthroplasty is in-joint, subluxed or dislocated, but this measure is too coarse to reliably estimate technical skill of the surgeon as fortunately acute dislocation is a rare complication of hemiarthroplasty⁴³⁸.

6.4.2 Feasibility assessment

I piloted taking these measurement on the hospital's Picture Archiving and Communication System (PACS) to check that it was feasible to do so. Dynamic Hip Screw patients do not routinely have post-operative radiographs taken, the final position of the implant is assessed at the end of the operation in the antero-posterior (AP) and lateral views using Image Intensification (II) before skin closure.

II images are not scaled by PACS, and so I manually scaled these (controlling for magnification) using a known constant implant dimension (screw barrel width).

6.4.3 Additional patient-based outcome measures for the trial

As discussed above, FPA seems like the most appropriate primary outcome measure for the trial, as surgical skill has the greatest bearing on the final product quality, before the many other factors that influence patient outcome during the course of post-operative recovery take effect (figure 21).

It seems sensible to consider other, non-FPA, non-radiograph based factors that are relevant to patient outcome and which may be influenced by the surgeon having received cadaveric simulation training. Given that this research is exploratory, it would be prudent to consider multiple ways in which cadaveric simulation training may benefit patients. These candidate secondary outcome measures will be discussed in turn.

6.4.4 Intra-operative blood loss

The volume of blood lost intra-operatively has direct clinical relevance, as hypovolaemia has negative physiological consequence⁴³⁹, particularly in the elderly with sub-optimal compensatory reserve⁴⁴⁰. Significant blood loss may necessitate blood transfusion, which carries risk to the patient⁴⁴¹ and a cost burden to the hospital⁴⁴².

Various patient-related factors influence intra-operative bleeding^{443 444}. These include pre-existing coagulopathies and recent/current use of anticoagulants. Procedural factors also affect the expected amount of surgical bleeding. An open surgical procedure through a large incision, involving dissection through vascular tissue, in a junctional area where a tourniquet cannot be applied, will naturally involve more blood loss than an arthroscopic or bloodless distal limb procedure with a tourniquet. Surgeon-factors also influence blood loss – particularly the choice of incision size/placement, and thoroughness of intra-operative haemostasis. This might include speed of recognition of bleeding, appropriate timely use of diathermy, and security of any haemostatic sutures. All of these surgeon-factors could be influenced by surgical skill, and it is reasonable to postulate that a less experienced surgeon might cause their patient to bleed more during their operation.

There are various ways to measure intra-operative blood loss⁴⁴⁵. The most accurate method for determining intra-operative blood loss in major open orthopaedic surgery is using the Hb-balance method described by Gao⁴⁴⁵. It is a scientifically logical method which is widely used in clinical studies, particularly in the arthroplasty literature⁴⁴⁵⁻⁴⁴⁷. Please see appendix 11.6 for details on how this was calculated.

6.4.5 Blood transfusion requirement

It follows that if a patient experiences significant intra-operative blood loss, they will be more likely to need a blood transfusion in the acute post-operative period (first 48

hours after surgery). This risk is also influenced by other variables such as pre-operative haemoglobin levels, coagulation status and cardiac reserve for tolerating the hyperdynamic circulatory physiology of anaemia. Blood transfusion is not without risk to patients⁴⁴¹, or cost to the NHS⁴⁴². It is therefore an important secondary outcome measure to record for DHS and hemiarthroplasty (for the non-surgeon reader; ankle fractures are fixed under tourniquet control so do not bleed very much).

6.4.6 Procedure Time

Procedure Time is also clinically relevant. The longer the operation, the longer the patient is under general anaesthesia, and the greater the risks from complications. This is particularly relevant to the hip fracture population, who tend to be frail, multi-morbid patients who are particularly vulnerable to the risks of prolonged anaesthesia⁴⁴⁸.

There is also an important health economic dimension to procedure time. Operating theatres are very expensive to run, the cost has been calculated at approximately £30 per minute. Therefore, even a relatively modest reduction in procedure time would have potentially significant cost savings for hospital trusts.

Procedure time has been shown in chapter 5 to demonstrate construct and concurrent validity in assessing surgical skill. I therefore think it is reasonable to use this as a secondary outcome measure for the trial.

6.4.7 Complication rate

For obvious reasons, complications represent a poor outcome for patients (and their surgeon). Complications might also mean a longer inpatient stay, more surgery or additional medical treatment. Given the imperative here to concentrate on events close to the time of surgery as outcome measures, to try and capture the impact of

surgical skill on patient outcome(s), I will focus on acute post-operative complications only. For the purposes of this thesis, I have defined 'acute' as a complication that occurs during the course of the admission. Another reason for this approach is that I can be sure of collecting good data on these, as they will be recorded in the discharge summary and coding data for the admission. A patient experiencing a late surgical complication such as DHS screw-cut out may re-present at a different hospital from where they had their original surgery, and hence these could be missed.

Complications will be considered in two categories. Medical complications will be considered as; hospital acquired pneumonia, renal complications (urosepsis, renal failure), cardiac complications (acute coronary syndrome), and inpatient death. Surgical complications will be considered as; wound complications (superficial wound infection/cellulitis, dehiscence, suture problems) and metalwork problems (deep infection, metalwork failure, re-fracture, dislocation etc.).

6.5 Conclusion

The radiographic parameters identified through this consensus setting exercise will form the core primary outcome set for measuring the technical skill of surgeon participants in the trial. These have been summarized in the CAD:TRAUMA radiology manual (please see appendix 11.5), which I will have as a companion guide in undertaking radiograph measurements.

6.6 Reflections

In this chapter's reflections, I will consider what I have learnt in undertaking the consensus setting exercise.

The work that forms the basis of this chapter was the first time I have built and administered a web-survey. I learnt a lot about how to build an online questionnaire, and through repeated beta-testing in Round 1, I quickly came to realise that it was more difficult than I first thought to achieve the balance between brevity and detail. Mindful of the fact that my sample group were all busy NHS consultants, I had deliberately kept the questions very short in round 1, but perhaps too much so as it was evident from the free text comments that some of them had misunderstood the objective of the exercise. I realised I was suffering the problem of being too close to the project to appreciate that those approaching it with fresh eyes and without a surgical education research interest may not realise that I was asking subtly different questions to 'what radiograph features predict clinical success'. I overcame this by piloting round 2 with some T&O colleagues to ensure the objectives were clear.

The other major lesson for me here was in how long it actually takes to do multiple survey rounds, as the enthusiasm of the participants naturally dwindled and people began to drop out after each round. The proportion of non-responders increased with each round, and the number who needed reminding increased. It was over 12 weeks between sending the first survey out and receiving the final responses in round 3.

I saw this as evidence that doctors as a group are survey fatigued, and if I were to do this again I would try and reduce the number of rounds and give tighter deadlines for completion, with more frequent reminders. I would also involve colleagues in piloting the survey before releasing it to be sure the objectives were clear.

Chapter 7: CAD:TRAUMA Study: Cadaveric simulation versus standard training for postgraduate trauma and orthopaedic surgical trainees; a multi-centre randomised controlled educational trial

In this chapter I will investigate whether an intensive cadaveric simulation training course for junior orthopaedic surgeons-in-training can lead to improved real world operative performance and improved patient outcomes, as compared to conventional 'on-the-job' training.

Declarations;

I received the following help in writing this chapter;

- Dr Nick Parsons advised on the statistical analysis design and checked it was conducted correctly

This chapter has been submitted for publication as a trial protocol;

James HK, Pattison GTR, Fisher JD, Griffin DR. Cadaveric simulation vs standard training for postgraduate trauma & orthopaedic surgical trainees: protocol for the CAD:TRAUMA study multi-centre randomized controlled educational trial.
Submitted to BMJ Open

Trial registration number: ISRCTN20431944

NHS research ethics approval: 15/WM/0464

Confidentiality Advisory Group approval: 16/CAG/0125

University Biomedical & Research Ethics Committee approval: REGO-2014-718

Sponsor: University of Warwick

7.1 Introduction

There is an urgent need to investigate alternatives to the current standard surgical training strategy of the master-apprentice model. For the reasons described in chapter 1, this model is no longer fit for purpose in a modern health service that is facing unprecedented challenges that impact both the quantity and quality of surgical training. There is great enthusiasm for cadaveric simulation training⁴⁴⁹ amongst the surgical education community, and increasing, albeit regionally inconsistent provision for T&O trainees within training programmes (please see national survey on simulation provision in chapter 4).

There is abundant low quality evidence of the positive educational impact of cadaveric simulation (please see chapter 3: systematic review on cadaveric simulation for postgraduate surgical training), but as yet no high quality evidence from a randomised trial showing improved patient outcomes from such training.

7.2 Aim

The aim of the trial described in this chapter is to determine which of two surgical training strategies leads to the best patient outcomes for three common orthopaedic procedures (dynamic hip screw, hemiarthroplasty and ankle fracture fixation) performed by junior surgeons-in-training.

7.3 Objectives

Primary Objective

- To assess the impact of a cadaveric simulation training intervention on patient outcomes following surgery performed by junior orthopaedic surgeons-in-training.

Secondary Objectives

- To define the early learning curve of three common trauma procedures in this group
- To explore the feasibility of using post-operative radiographs to assess technical skill.

7.4 Methods

This study was conducted in line with Medical Research Council Good Clinical Practice principles and guidelines and Warwick Clinical Trials Unit standard operating and governance procedures. The trial has been reported in accordance with the Consolidated Standards of Reporting Trials (CONSORT) guidelines⁴⁵⁰.

7.4.1 Trial design

A randomised controlled trial of cadaveric simulation training versus standard training for postgraduate trauma and orthopaedic surgical trainees.

7.4.2 Sample size

This is the first attempt to objectively measure transfer of open surgical skills from the cadaveric simulation laboratory into the workplace using patient outcome measures. I decided to take a pragmatic approach to sample size given the exploratory nature of the study. The surgical training centre can accommodate 16 delegates at one time, and financial resources permitted one iteration of the course. The maximum sample size was therefore 16 participants in each arm of the study.

7.4.3 Participants

Inclusion criteria;

- 1) Trauma & Orthopaedic surgeon-in-training
- 2) In-post on August 5th 2014*
- 3) Specialty training year 1, 2 or 3 at recruitment.

Exclusion criteria;

- 1) Consent refusal
- 2) Unavailability on either of the dates of the cadaveric simulation training courses.

*This date was chosen as it is the start of the training year.

7.4.4 Interventions

Cadaveric Simulation Training

The intervention was an intensive, two-day cadaveric simulation training course held at the West Midlands Surgical Training Centre (WMSTC) (figure 28), where four trauma procedures were taught; 1) Dynamic hip screw and 2) hemiarthroplasty for fractured neck of femur, 3) ankle fracture fixation and 4) lower limb fasciotomy. Procedures 1-3 were chosen because they are very common procedures that are often performed by junior trainees, and form the bulk of the operating exposure this stage of surgeon will encounter in the workplace.

These procedures facilitate the development of surgical skills that can then be further developed for more sophisticated procedures further on in training, for example the cemented hemiarthroplasty operation shares many steps in common with total hip arthroplasty (hip replacement). They are mapped to the curriculum as competency is required to be demonstrated by ST3 level, and are known as ‘index’ procedures⁴⁵¹.

Fasciotomy was included as a ‘filler’ to make the course structure work as an even number of procedures was needed to have two parallel circuits of four stations. This procedure was chosen as it is an important high-stakes, anatomically-based operation that is rarely performed by trainees. Outcomes related to the fasciotomy procedure were not collected or included in the analysis. The course was delivered in early September 2014, which represents the beginning of the surgical training year, which runs from August-August.

Standard training

The control group received the current standard ‘on-the-job’ training under the master-apprentice model.

7.4.5 Follow-up

Study participants were followed up for the 2014/15 training year. The control group received the course at the end of the follow-up period, as a condition of ethical approval for the study, to ensure there was equity of training opportunity amongst study participants. A schematic overview of the study timeline is shown in figure 24 below.

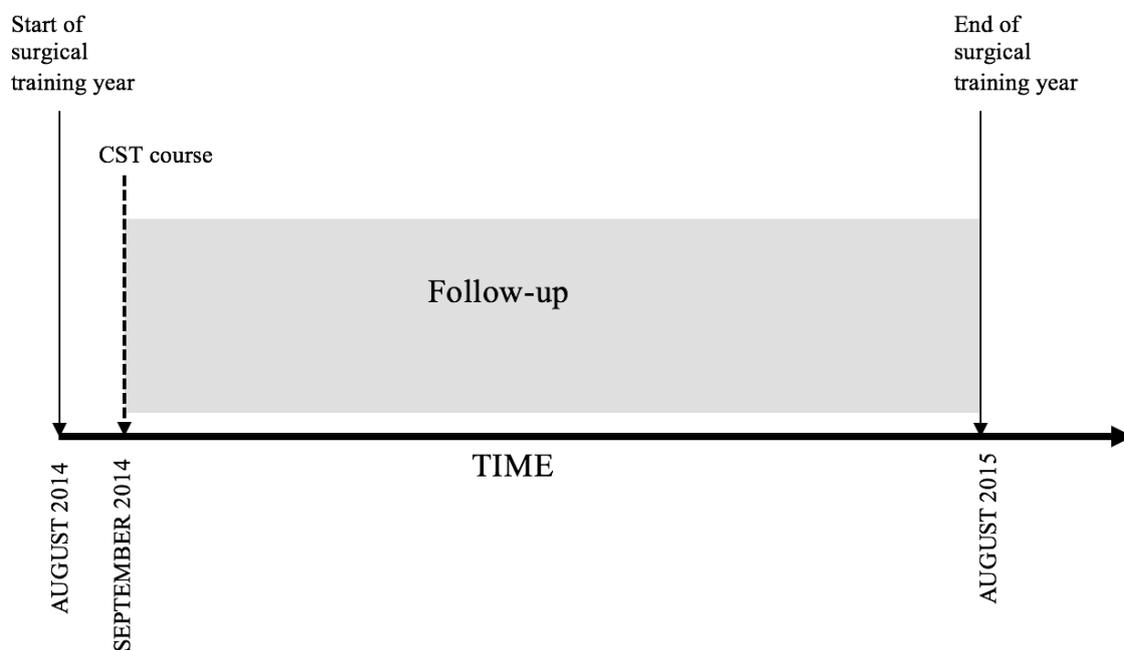


Figure 24. Schematic overview of the study timeline

7.4.6 Data Collection

I used radiographic and clinical outcome measures to assess the impact of the training. I identified the operations that had been performed by the participants during study follow up by accessing their electronic surgical logbooks ('eLogbook')⁴⁵². This data was extracted centrally for me by the eLogbook data team. This gave me the date, location and hospital patient identifying numbers so that I could locate the relevant radiographs and clinical information for each case. I also collected supervision codes so that only operations that were coded as 'supervised-trainer scrubbed' or 'supervised-trainer unscrubbed' were included in the analysis (those coded as 'assisted' are done by the consultant and were excluded).

1) Radiographic outcomes

Intra-operative (DHS) and post-operative radiographs (hemiarthroplasty, ankle fracture fixation) were obtained from hospital electronic servers. I measured the implant position using the manual measurement tools in the hospital Picture Archive and Communication System (PACS). The measurements vary by operation type and were defined following the Delphi exercise described in chapter 6. Please see the CAD:TRAUMA radiology manual in appendix 11.5 for detailed description, diagrams and supporting evidence for the measurements.

Dynamic Hip Screw

A. Primary Outcome

- Tip-Apex distance (in mm)

B. Secondary Outcomes (in order of importance)

- Lag screw position in the femoral head (defined by Cleveland Zones)
- Plate flush to the lateral femoral cortex (Yes or No)
- 8 cortex hold for plate screws (Yes or No)

Hemiarthroplasty

A. Primary Outcome

- Leg length discrepancy (mm)

B. Secondary Outcomes (in order of importance)

- Femoral stem alignment (degrees off neutral)
- Cement mantle quality (Barrack grade score)
- Femoral offset change relative to native hip (mm)

Ankle fracture fixation

A. Primary Outcome

- Medial clear space (mm)

B. Secondary Outcome

- Lateral malleolar displacement (mm)
- Tibiofibular clear space (mm)
- Talocrural angle (degrees)
- Medial Malleolar displacement (mm)

2) Non-radiographic outcomes

The non-radiographic outcome measures were 1) procedure time, 2) intra-operative radiation dose to the patient (DHS/ankle fracture fixation), 3) Intra-operative blood loss (hemiarthroplasty), 4) post-operative complication rate (DHS/hemiarthroplasty) and 5) mortality at 12 months (DHS/hemiarthroplasty). These were obtained from the hospital databases at each site. Procedure time was obtained separately from the theatre management system. Details about complications were obtained from the discharge summaries and clinic letters (where relevant).

- Procedure Time

Procedure time was defined as knife-to-skin/surgical start time to wound closure/surgical stop time. I chose this as an outcome measure as there is evidence in the literature that procedure time is inversely related to technical proficiency. Please see Chapter 5, Table 10 section 4 for details of evidence.

- Intra-operative radiation dose to patient

Image intensification (II) is routinely used intra-operatively in DHS and ankle fracture procedures, so that the position of the fracture and implant(s) can be adjusted and optimized. II use was defined in the study as ‘time under fluoroscopy’ (in seconds) and radiation dose (in mGym²). There is evidence showing that surgeons need to use less II intra-operatively with increasing skill (Please see Chapter 5, Table 10 section for further details of evidence). Hemiarthroplasty does not require II and so this outcome measure was not used here.

- Intra-operative blood loss

It seemed reasonable to measure the intra-operative blood loss for hemiarthroplasty, as logically it could be that a more technically proficient surgeon might cause less bleeding and perform better haemostasis. I calculated intra-operative blood loss using the Haemoglobin balance method, as this has been shown elsewhere to be the most reliable method of estimating blood loss after arthroplasty⁴⁴⁵. Please see appendix 11.6 for a detailed description of this method.

- Post-operative complication rate

Post-operative complications obviously represent a poor outcome for the patient. I focused on acute post-operative complications during the inpatient admission. These are sub-categorised as acute medical complications (hospital acquired pneumonia, renal complications, cardiac complications, thromboembolic events; deep vein

thrombosis or pulmonary embolus, death in hospital) and surgical complications (wound infection, wound dehiscence, metalwork failure, deep infection).

- Mortality at 12 months

It is known that mortality at 12 months following hip fracture surgery is significant⁴⁵³. This likely reflects the frailty of the hip fracture population and the often pre-terminal nature of the injury. Mortality data were not collected for ankle fracture patients as these are generally a younger and fitter group.

7.4.7 Sampling

The West Midlands deanery oversees three Trauma & Orthopaedic Specialist Training Programmes; Warwick, Oswestry and Birmingham, totalling approximately 112 trainees. Please see Chapter 4 Figure 12 for detailed description of training territories. Sampling was restricted to within the West Midlands Deanery to reduce the geographic variability in the training programme structure which could confound the study results.

7.4.8 Recruitment

Eligible surgeons-in-training were identified by liaison with the West Midlands Deanery training programme administrators. Support for the study was agreed prospectively with the Training Programme Directors in the respective three schools. An invitation email was sent to the eligible trainees, and I displayed study recruitment posters in doctor's offices and ward/theatre coffee rooms.

7.4.9 Randomisation

At recruitment, participants were randomly allocated to either the intervention or the control groups on a 1:1 basis. The allocation sequence was generated prior to recruitment using a simple blocking scheme (block size = 4), to limit the possibilities of imbalance occurring at any stage during the course of the study.

7.4.10 Consent

Surgeon participants:

Study participants were given written and verbal information about the study, and I obtained written consent from all participants. Consent forms were signed by participants and held on file. The right to refuse participation without giving reasons was fully respected, as was the right to withdrawal from the study at any time without prejudice to their training.

Patients whose operations were assessed:

Patients who underwent an operation by a surgeon in the study were not separately consented to allow access to their radiographs or clinical outcome data. This decision was taken in agreement with project supervisors that seeking consent from a group of primarily elderly, frail patients to assess low risk, routine clinical data in a secure manner for a trial they are not directly participating in would be unduly burdensome for the patients. Permission to access this information for the study without patient consent was granted by the Confidentiality Advisory Group (16/CAG/0125).

7.4.11 Statistical methods

The study is reported following CONSORT guidelines⁴⁵⁰. Baseline demographic data are compared between the two arms of the study. A CONSORT flow chart is used to show participant flow through the study (figure 31). Results of analyses for the three procedures (DHS, hemiarthroplasty, ankle fracture fixation) are reported separately.

The main focus of the analysis is to report differences between the two study groups (intervention and control) with respect to the implant positions as measured from radiographs and the clinical outcome measures detailed above.

Statistical tests were two-sided and considered to demonstrate a significant difference when $p < 0.05$. My analysis has followed three approaches;

1. Temporal trends by group allocation for implant position, procedure time and radiation dose (where relevant) are presented, to give a visual impression of change in these parameters over time, by study group.
2. Linear mixed effects regression models were fitted to data to allow for within-surgeon correlation between repeated observations (surgeon clustering as a random effect). The regression models also adjusted important covariates such as patient condition (measured by American Society of Anaesthesiologists grade⁴⁵⁴), patient age and surgeon experience. The results of model fitting are summarized by plotting individual learning curves, and overall (intervention and control group) learning curves for the two arms of the study.
3. Basic descriptive analysis of between group comparisons for binary radiographic measures, complications and mortality is presented. In this analysis, I have considered the operations as independent variables and ignored autocorrelation within individual surgeons and temporal effects of

the training. I thought it was reasonable to present it in this way as it is much simpler and clearer to describe and discuss, and within my remit of ability as a non-expert statistician. This approach is justified statistically based on the fact that the temporal distribution of the observations seems to be well balanced between both groups (i.e. there does not appear to be any significant temporal clustering when looking at the scatterplots of observations by time), and natural between-surgeon variations in ability should be balanced between the groups by randomisation (baseline demographic data show the groups are comparable at baseline in age and experience levels (table 12, below)).

All analysis was undertaken in SPSS (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp).

7.4.12 The CST training intervention

Course Design

The CST course ran on 25-26th September 2014, at the beginning of the surgical training year. The course took place at the West Midlands Surgical Training Centre (WMSTC) at the University Hospital Coventry & Warwickshire (UHCW). The WMSTC is a specialized wet-laboratory facility for delivering cadaveric training, and has an experienced dedicated faculty to facilitate training delivery.

The course ran for two full days, with 9 expert consultant faculty teaching on eight fresh-frozen hemi-cadavers (waist-to-toe-tip). The cadaveric specimens were purchased under license from a specialist supplier⁴⁵⁵ as there is not yet a functional local body donation programme at UHCW.

The participant: faculty ratio was 2:1, and participant: cadaver ratio was 2:1.

Course Delivery

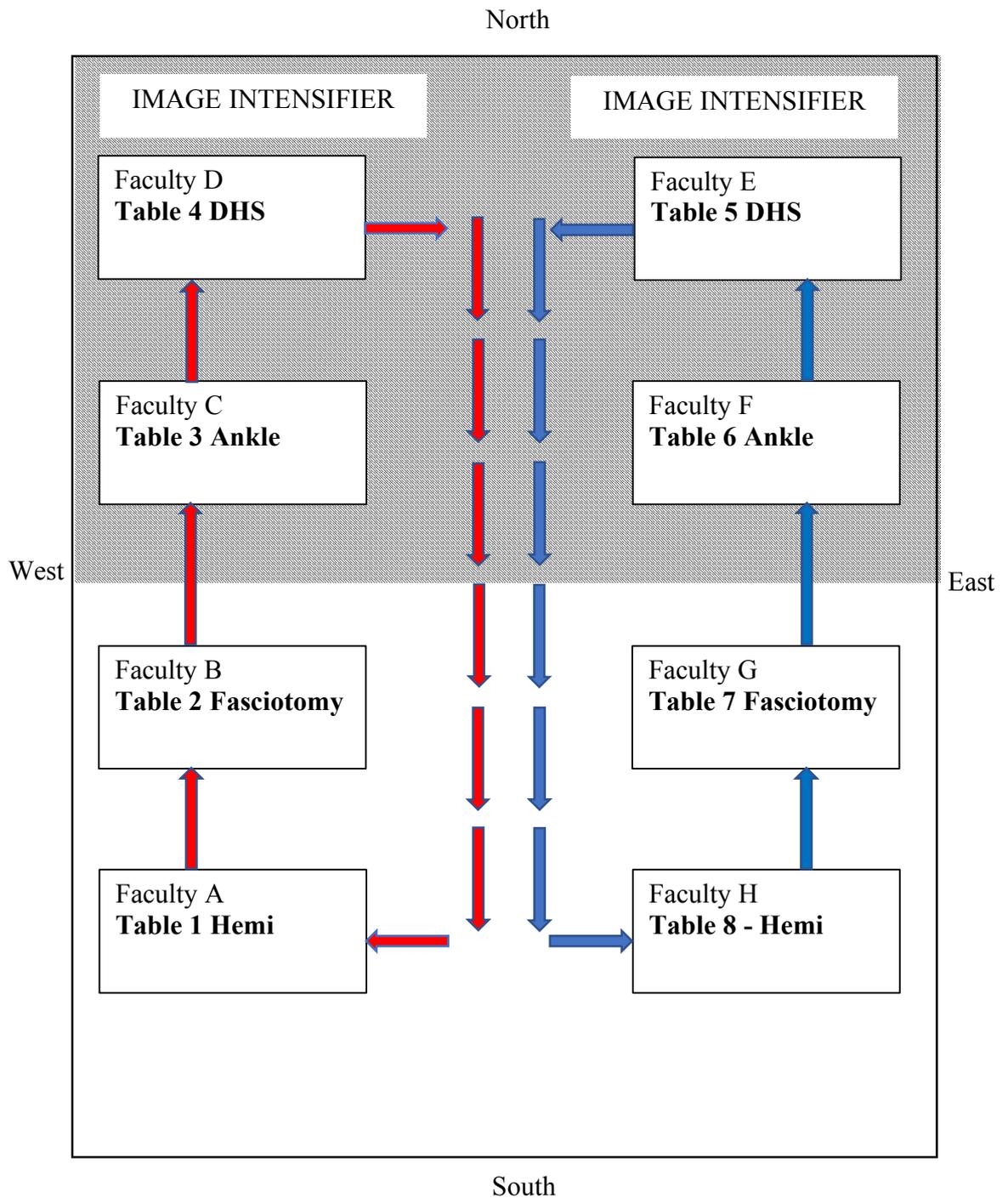
The cadaveric specimens were set up on individual tables within the WMSTC, with each table functioning as an operating theatre. The tables each have their own operating light system overhead.

I endeavoured to maximize the physical, environmental and psychological fidelity of simulation by the following steps;

- Full surgical dress including masks, gloves, gowns and lead x-ray aprons were worn by participants
- The usual disposable surgical drapes were used
- Skin preparation (iodine solution) was used by participants to prepare the surgical site, and participants and faculty were asked to observe the usual sterile field precautions as in real theatre
- Full surgical instrument trays, surgical implants and cement (for hemiarthroplasty) of the same type as in real theatre was used
- Image intensifier (mobile x-ray) was available for intra-operative use
- Background noise levels and room temperature were maintained at what would usually be expected in the operating theatre
- Scrub staff were assigned to the stations to behave as they usually would in the operating theatre and medical students were asked to act as anaesthetists. The study participants were paired on each station, taking alternate turns to be first surgeon and then assistant. There was therefore a sufficient volume of people around each station such that the study participants felt suitably scrutinized and crowded as happens in the real operating theatre.
- Faculty were asked to prompt the trainees appropriately as they would in real life theatre, with the exception that they could allow trainees to make mistakes or take longer with procedural steps than they would in real-life. A trainee could be allowed to struggle with the procedure and decide on the resolution themselves in a way that would be unsafe with real patients.

The 8 simulated operating theatres were set up within the WMSTC as two parallel round-robin circuits of 4 (figure 25). The two stations requiring x-ray use (DHS and ankle ORIF) were set up at the far end of the room to create a radiation zone where appropriate precautions were used, per guidelines^{456 457}. Careful consideration was given to the optimum sequential use of the cadaveric specimens in planning the course structure. For example, it was necessary that the DHS station preceded the hemiarthroplasty station as it would obviously not be possible to perform a DHS operation when the femoral head had been removed. Similarly, the fasciotomy incisions would compromise the soft tissue envelope of the lower limb to a sufficient degree that the fidelity of the ankle ORIF station would be compromised. Left sided procedures were performed on day 1, and right sided procedures on day 2.

It was necessary to make the best and most efficient use of the cadaveric material, for both moral and financial reasons.



Notes:

Hatched area = radiation control zone

Arrows = blue or red stream

Surgeons, scrub staff and trolleys are positioned south of the tables

Cadaveric feet point west

All surgery on left side (Day 1) and right side (Day 2)

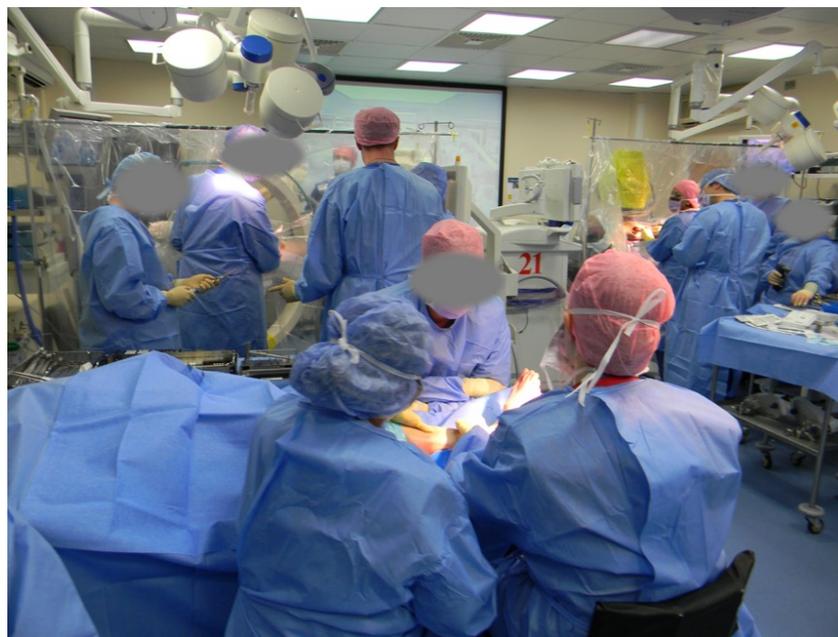
Figure 25. Schematic overview of course set-up in the surgical training centre

Time	Activity	Role	Table 4 DHS	Table 3 ANKLE	Table 2 FASCI	Table 1 HEMI	Lecture theatre
RED STREAM							
0800	Day Briefing						All
							Study brief
0830	Session 1	Surgeon	Trainee 1	Trainee 3	Trainee 5	Vacant	Trainee 7
		Assistant	Trainee 2	Trainee 4	Trainee 6	Vacant	Trainee 8
1000	Turnaround						
1015	Coffee						
1030	Session 2	Surgeon	Trainee 4	Trainee 6	Trainee 8	Trainee 2	Vacant
		Assistant	Trainee 3	Trainee 5	Trainee 7	Trainee 1	
1200	Turnaround						
1215	Lunch						
1300	Session 3	Surgeon	Trainee 5	Trainee 7	Trainee 1	Trainee 3	Vacant
		Assistant	Trainee 6	Trainee 8	Trainee 2	Trainee 4	
1430	Turnaround						
1445	Tea						
1500	Session 4	Surgeon	Trainee 8	Trainee 2	Trainee 4	Trainee 6	Vacant
		Assistant	Trainee 7	Trainee 1	Trainee 3	Trainee 5	
1630	Turnaround						
1645	Session 5	Surgeon	Vacant	Vacant	Vacant	Trainee 7	Study brief
		Assistant	Vacant	Vacant	Vacant	Trainee 8	Trainee 1-2
							Trainee 3-4
							Trainee 5-6
1815	Debrief						All
1830	Close						
Time	Activity	Role	Table 5 DHS	Table 6 ANKLE	Table 7 FASCI	Table 8 HEMI	Lecture theatre
BLUE STREAM							
0800	Day Briefing						All
							Study brief
0830	Session 1	Surgeon	Trainee 9	Trainee 11	Trainee 13	Vacant	Trainee 15
		Assistant	Trainee 10	Trainee 12	Trainee 14	Vacant	Trainee 16
1000	Turnaround						
1015	Coffee						
1030	Session 2	Surgeon	Trainee 12	Trainee 14	Trainee 16	Trainee 10	Vacant
		Assistant	Trainee 11	Trainee 13	Trainee 15	Trainee 9	
1200	Turnaround						
1215	Lunch						
1300	Session 3	Surgeon	Trainee 13	Trainee 15	Trainee 9	Trainee 11	Vacant
		Assistant	Trainee 14	Trainee 16	Trainee 10	Trainee 12	
1430	Turnaround						
1445	Tea						
1500	Session 4	Surgeon	Trainee 16	Trainee 10	Trainee 12	Trainee 14	Vacant
		Assistant	Trainee 15	Trainee 9	Trainee 11	Trainee 13	
1630	Turnaround						
1645	Session 5	Surgeon	Vacant	Vacant	Vacant	Trainee 15	Study brief
		Assistant	Vacant	Vacant	Vacant	Trainee 16	Trainee 9-10
							Trainee 11-12
							Trainee 13-14
1815	Debrief						All

Figure 26. Course timetable – Day 1: left sided procedures

Time	Activity	Role	Table 4 DHS	Table 3 ANKLE	Table 2 FASCI	Table 1 HEMI	Lecture theatre
RED STREAM							
0800	Day Briefing						All
0830	Session 1	Surgeon	Trainee 10	Trainee 12	Trainee 14	Vacant	Vacant
		Assistant	Trainee 9	Trainee 11	Trainee 13	Vacant	
1000	Turnaround						
1015	Coffee						
1030	Session 2	Surgeon	Trainee 11	Trainee 13	Trainee 15	Trainee 9	Vacant
		Assistant	Trainee 12	Trainee 14	Trainee 16	Trainee 10	
1200	Turnaround						
1215	Lunch						
1300	Session 3	Surgeon	Trainee 14	Trainee 16	Trainee 10	Trainee 12	Vacant
		Assistant	Trainee 13	Trainee 15	Trainee 9	Trainee 11	
1430	Turnaround						
1445	Tea						
1500	Session 4	Surgeon	Trainee 15	Trainee 9	Trainee 11	Trainee 13	Vacant
		Assistant	Trainee 16	Trainee 10	Trainee 12	Trainee 14	
1630	Turnaround						
1645	Session 5	Surgeon	Vacant	Vacant	Vacant	Trainee 16	Debrief
		Assistant	Vacant	Vacant	Vacant	Trainee 15	Trainee 9-10
							Trainee 11-12
							Trainee 13-14
1815	Debrief						Trainee 15-16
1830	Close						
Time	Activity	Role	Table 5 DHS	Table 6 ANKLE	Table 7 FASCI	Table 8 HEMI	Lecture theatre
BLUE STREAM							
0800	Day Briefing						All
0830	Session 1	Surgeon	Trainee 2	Trainee 4	Trainee 6	Vacant	Vacant
		Assistant	Trainee 1	Trainee 3	Trainee 5	Vacant	
1000	Turnaround						
1015	Coffee						
1030	Session 2	Surgeon	Trainee 3	Trainee 5	Trainee 7	Trainee 1	Vacant
		Assistant	Trainee 4	Trainee 6	Trainee 8	Trainee 2	
1200	Turnaround						
1215	Lunch						
1300	Session 3	Surgeon	Trainee 6	Trainee 8	Trainee 2	Trainee 4	Vacant
		Assistant	Trainee 5	Trainee 7	Trainee 1	Trainee 3	
1430	Turnaround						
1445	Tea						
1500	Session 4	Surgeon	Trainee 7	Trainee 1	Trainee 3	Trainee 5	Vacant
		Assistant	Trainee 8	Trainee 2	Trainee 4	Trainee 6	
1630	Turnaround						
1645	Session 5	Surgeon	Vacant	Vacant	Vacant	Trainee 8	Debrief
		Assistant	Vacant	Vacant	Vacant	Trainee 7	Trainee 1-2
							Trainee 3-4
							Trainee 5-6
1815	Debrief						Trainee 7-8
1830	Close						

Figure 27: Course timetable – Day 2: right sided procedures



Top left: Figure 28. The West Midlands Surgical Training Centre

Above: Figure 29. A cadaveric DHS being performed

Bottom left: Figure 30. A cadaveric ankle fracture fixation being performed during the course (two DHS simulations in background)

7.5 Results

7.5.1 Flow of participants

Forty individuals were assessed for eligibility and invited to participate in the study. Thirty-three of these were randomized. Of the seven who were not randomised, four did not respond to the invitation and three declined to participate. Eighteen participants were randomised to the intervention group, and fifteen to the control group. Eleven of eighteen intervention group participants received the allocated intervention. Five participants withdrew from the study post-randomisation; one due to illness, one due to a career change, and three for undisclosed reasons. Two participants were subsequently unable to attend the training course in September due to clinical commitments and were swapped to the control group, and considered protocol violators. Ten of fifteen participants allocated to the control group received standard training. Two participants allocated to the control group withdrew from the study post-randomisation, for undisclosed reasons. Three participants in the control group were unable to attend the training course offered at the end of study follow-up due to clinical commitments and so were swapped to the intervention group, and considered protocol violators. There were therefore 14 participants in group 1 (intervention group), and 12 participants in group 2 (control group). All participants completed follow-up. The flow of participants through the study is shown in the CONSORT diagram (figure 31).

A per-protocol analysis ('as trained') will be presented as the primary analysis, with an intention-to-treat ('as randomised') analysis presented as a sensitivity analysis (appendix 11.7).

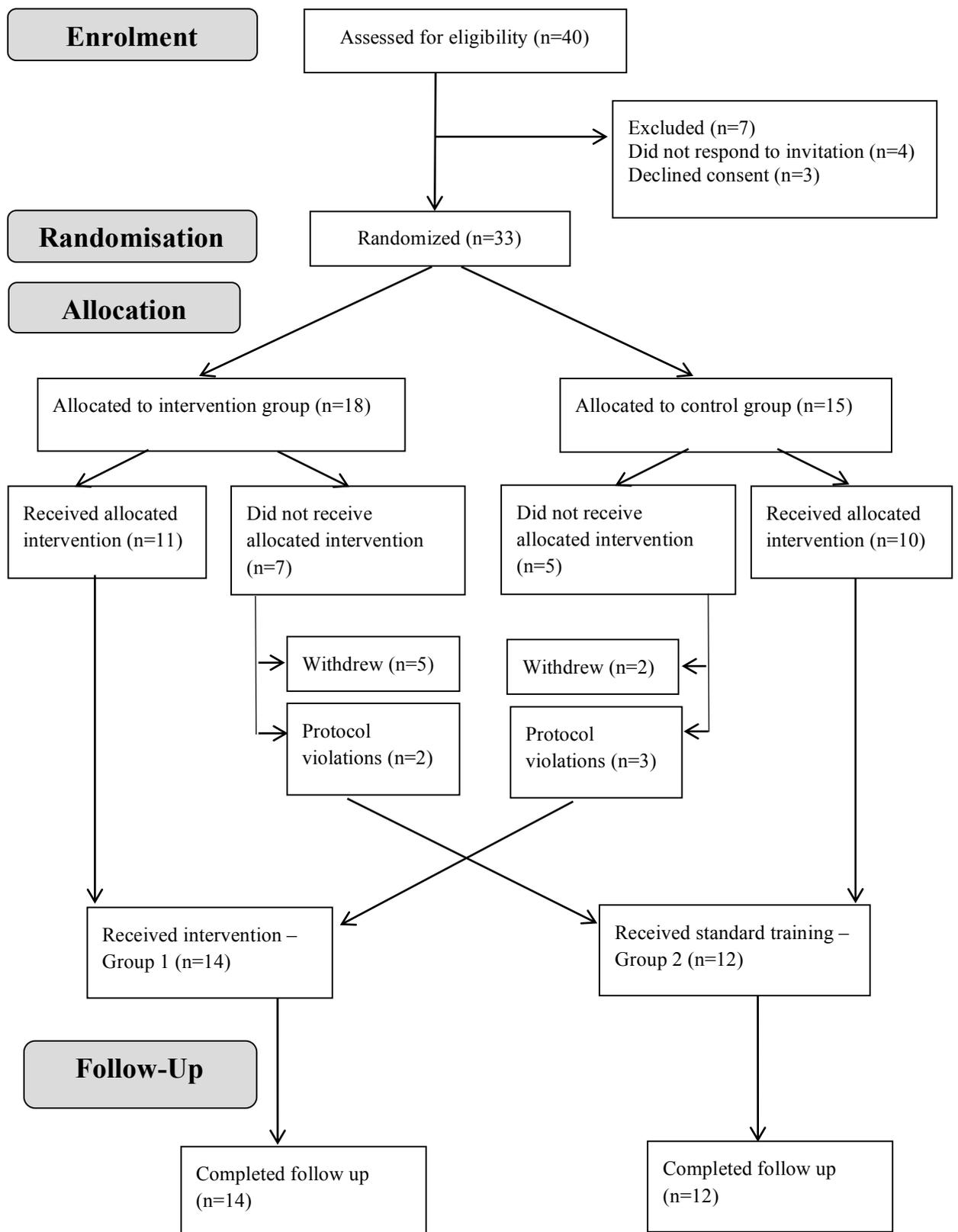


Figure 31. CONSORT diagram of participant flow

7.5.2 Participant demographics

The baseline demographics of the participants are shown in table 12.

Table 12. Baseline participant characteristics

	Intervention Group (n=14)	Control Group (n=12)
Age (years)		
Mean (SD)	28 (1.73)	30 (5.01)
Range	25-31	26-37
Gender		
Male (%)	9 (64)	10 (83)
Female (%)	5 (36)	2 (17)
Completed months T&O training at baseline		
Mean (SD)	12 (11.73)	17 (14.23)
Range	0-34	3-54
Baseline operative experience*		
DHS		
Mean (SD)	14 (9.39)	10 (9.57)
Range	0-33	0-24
Hemiarthroplasty		
Mean (SD)	9 (9.65)	8 (10.26)
Range	0-27	0-29
Ankle ORIF		
Mean (SD)	6 (5.58)	4 (4.71)
Range	0-17	0-15

*Number of cases performed at supervision levels STS/STU/P

The age of participants is similar between the two groups. There were relatively more female participants in the intervention group than in the control group (36%, compared to 17% in the control group). The participants in the intervention group had completed less months of T&O training prior to the study than had the control group (mean 12 months for the intervention group, vs 17 months for the control group). The intervention group had performed on average 4 more DHS cases, 1 more hemiarthroplasty case and 2 more ankle ORIF cases prior to the study as compared

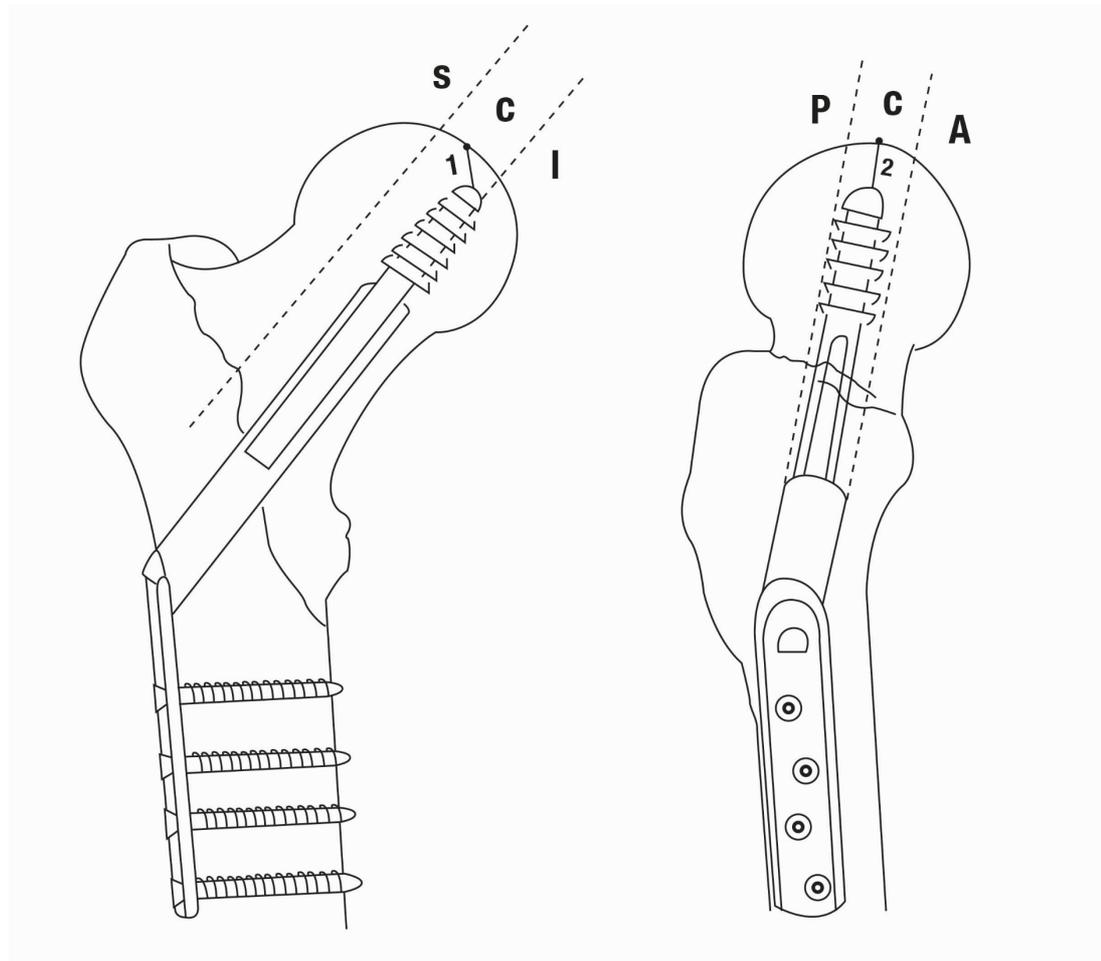
to the control group, but the overall range of experience for these cases between the two groups was comparable.

7.5.3 Outcomes of surgery

1. Dynamic Hip Screw

There were 317 DHS operations performed by the study participants during follow up, 196 by the intervention group and 121 by the control group. Two-view radiographs were available for 174 and 114 cases respectively.

A schematic overview of radiographic measurements for DHS is presented for clarity.



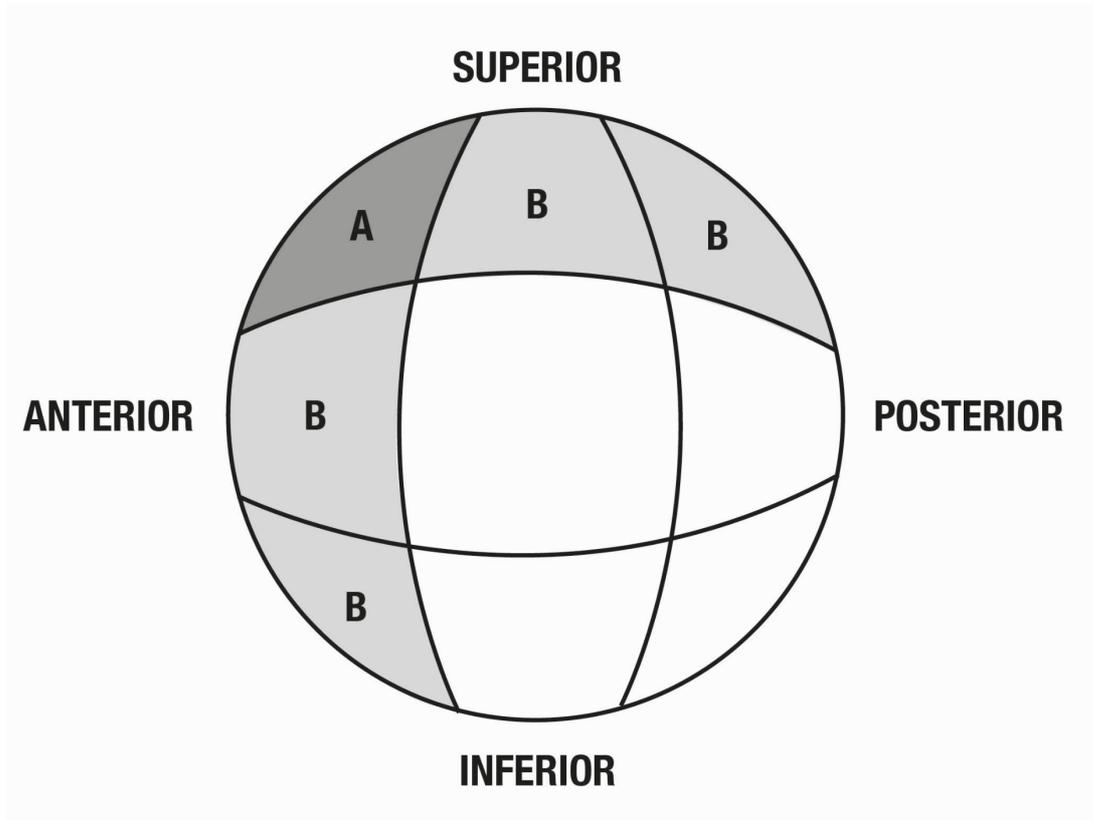
Anteroposterior (AP) view

Lateral view

Tip Apex Distance (TAD) = Distance 1(mm) + Distance 2(mm)

Femoral head zones:	S = Superior	}	AP view
	C = Central		
	I = Inferior		
	P = Posterior	}	Lateral view
	C = Central		
	A = Anterior		

Figure 32. Diagram of DHS in-situ showing Tip-Apex Distance and femoral head position (adapted from the AO Foundation⁴¹²)



View looking directly at the femoral head

Zone A = anterior **and** superior position

Zone B = anterior **or** superior position

Fig 33. Modified Cleveland's femoral head zones

Primary Outcome

The primary outcome measure is the quality of operation as measured by Tip-Apex Distance.

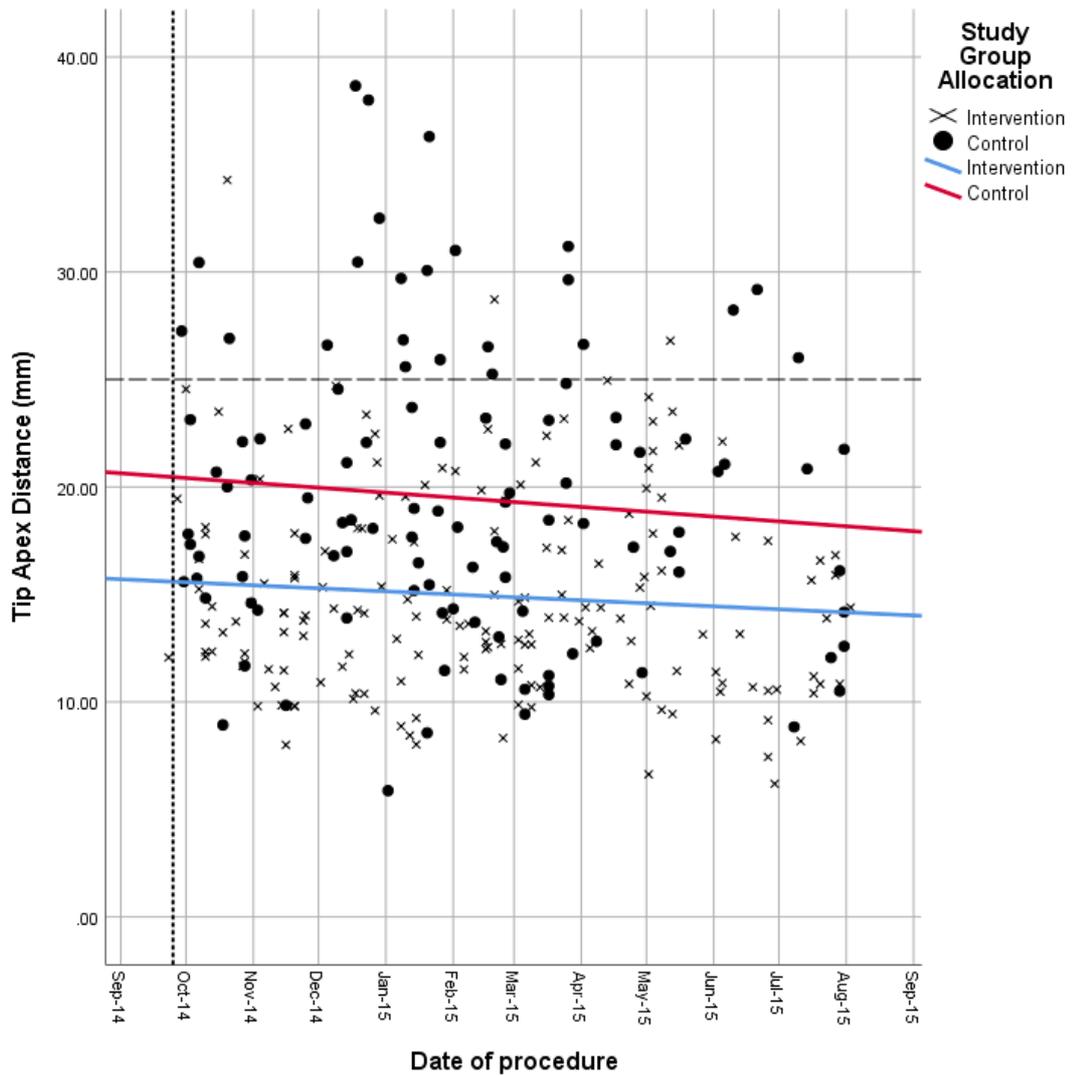


Figure 34. Tip-Apex Distance by study group

A TAD > 25mm predicts a poor clinical outcome, due to increased cut-out rate ^{414 458}. This is shown on the graph above as a reference line on the y-axis.

A comparison of ‘poor’ TAD (i.e. >25mm) between groups shows that there were significantly more poorly positioned implants in the control group; 1.7% (3 of 172) implants in the intervention group were poorly positioned vs 20.7% (23 of 111) in the control group, $p < 0.001$ (χ^2 test).

A DHS screw that is superiorly positioned in the head in the antero-posterior (AP) view, and/or anteriorly positioned in the head in the lateral view (a ‘high +/- anterior screw’) is significantly more likely to fail than a screw that is placed centrally or inferiorly⁴⁵⁹⁻⁴⁶¹. This risk of cut-out is independent of TAD⁴⁵⁹. A comparison of both groups shows that there were significantly more high +/- anterior screws in the control group, compared to the intervention group; 17% of screws were high +/- anterior in the intervention group vs 37% in the control group, $p < 0.001$ using a chi squared test.

Table 13. Tip-Apex distance and lag screw zone position by study group

	Number of cases (%)	
	Intervention Group	Control Group
Tip-Apex Distance		
≤25mm	169 (98.3)	88 (79.3)
>25mm	3 (1.7)	23 (20.7)
Total	172*	111*
Difference (χ^2) p=<0.001 (OR=14.7, 95%CI 4.30-50.39)		
Position of lag screw in femoral head		
Central +/- inferior	145 (83)	72 (63)
Superior or anterior	27 (16)	41 (36)
Superior and anterior	2 (1)	1 (1)
Total	174	114
Difference (χ^2) p=<0.001 (OR 2.92, 95%CI=1.68-5.06)		

* 3 cases in the control group and 2 in the intervention group had inadequate images of the apex of the femoral head in one or both views which precluded accurate measurement of TAD.

Learning curve analysis shows there is a significant difference between the learning curves for the two arms of the study, with the intervention group improving more quickly with respect to implant position (TAD) than the control group. Adding the information about the group allocation to the linear mixed effects model significantly improved the overall fit of the model, suggesting there is a true difference in the shapes of the learning curves (Likelihood ratio test (F-test), $p= 0.001$).

Secondary Outcomes

i. Radiographic secondary outcome measures

Table 14. DHS radiographic secondary outcome measures by group

	Intervention	Control
Plate flush to lateral cortex (cases)		
Yes	119	62
No	59	52
Total cases	178	114
Difference (χ^2) p= 0.032 (OR =1.69, 95%CI 1.04-2.74)		
8 cortex screw hold (cases)		
Yes	175	109
No	3	3
Total cases	178	112*
Difference (χ^2) p=0.563 (OR1.61, 95%CI 0.318-8.10)		

* Full length of plate could not be seen for two cases

There were significantly more plates that were not flush to the lateral femoral cortex in the control group ($p=0.032$, χ^2 test). There is no difference between groups for 8-cortex screw hold.

ii. Non-radiographic secondary outcome measures

Procedure Time

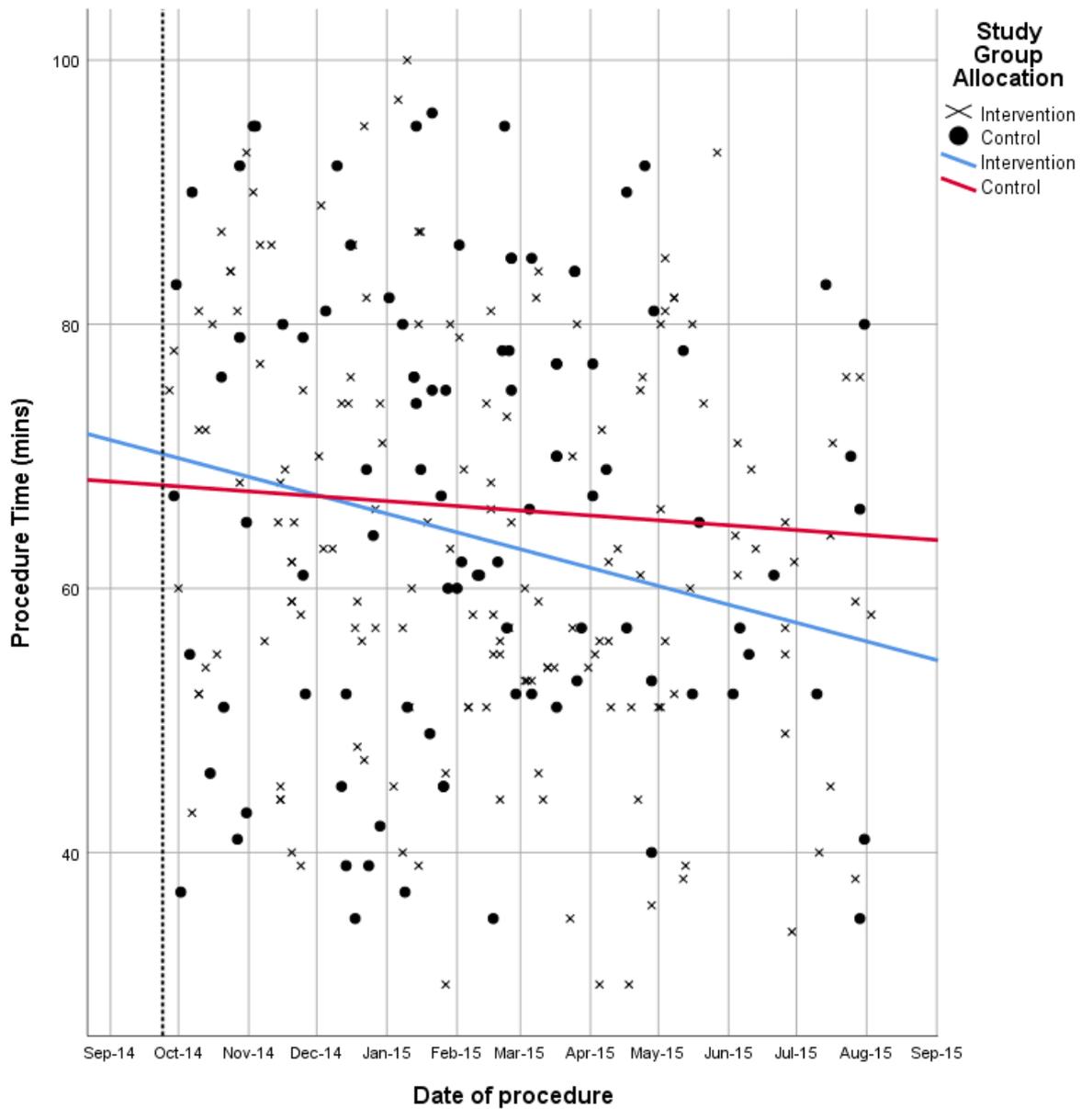


Figure 35. Procedure time by study group

Mean procedure time in the intervention group was 63.8 minutes (range 30-100 minutes) and was 66.1 minutes (range 30-100) in the control group. The difference in mean procedure time between groups was not significant using a t-test of independent samples ($p=0.271$, mean difference = -2.29, 95% CI -6.38 to -1.780)

Learning curve analysis

I fitted the same linear mixed effects model described above to see if there was a difference between the learning curve in performance time improvement between the two groups. This additional analysis suggests there was no significant difference in the rate of improvement between groups (likelihood ratio test (F-test), $p=0.193$)

Intra-operative radiation dose to patient

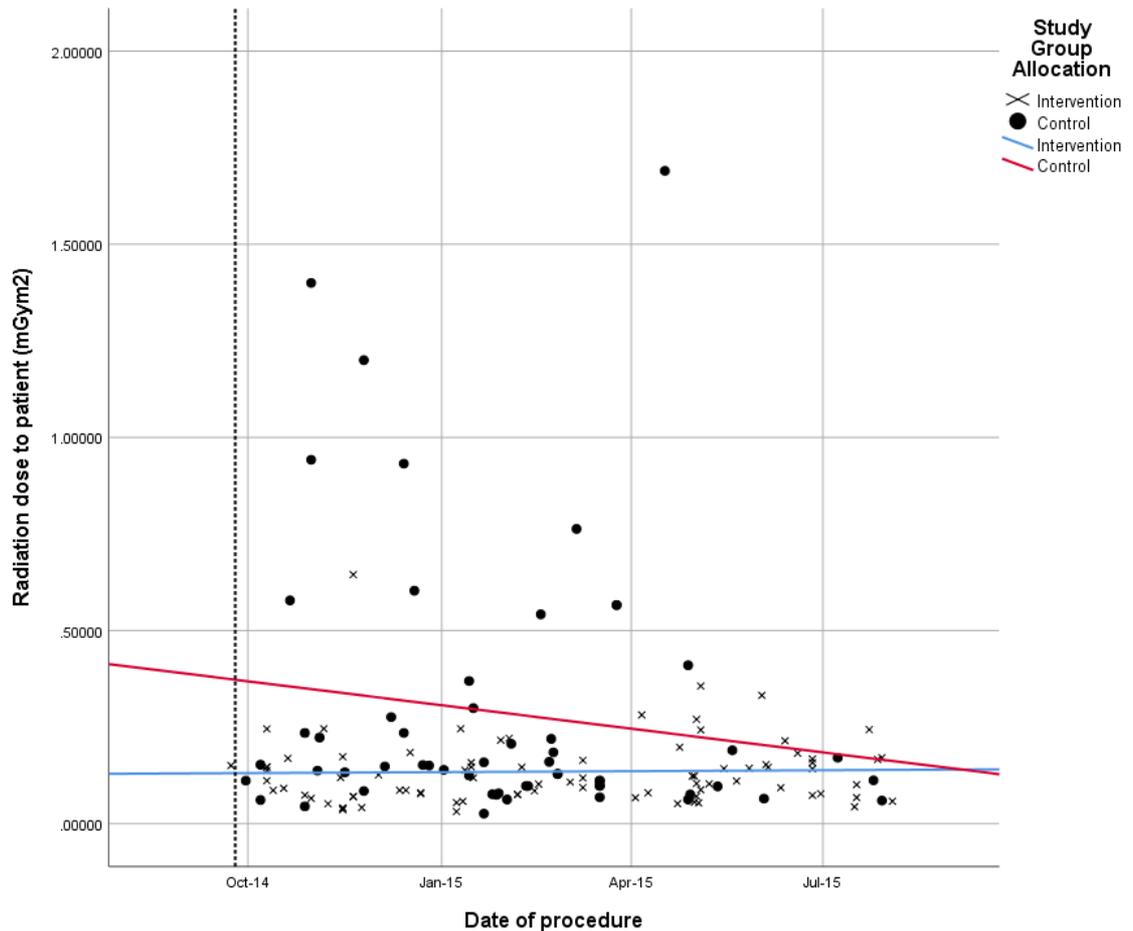


Figure 36. Intra-operative radiation dose to patient

Surgeons from the intervention group used significantly less radiation on average during procedures compared to the control group surgeons ($p < 0.001$, independent samples t-test). The mean radiation dose to patients in the intervention group was 0.134mGym^2 (range 0.031-0.644, $\text{SD}=0.88$) and 0.286mGym^2 (range 0.026-1.690, $\text{SD}=0.35$) in the control groups.

Learning curve analysis shows no statistical difference between the rate of improvement between the groups ($p=0.542$, likelihood ratio test (F-test)).

Post-operative complication rate

Table 15. Number of DHS patients experiencing complications

	Intervention Group	Control Group
Number of patients with complications;		
1. Hospital Acquired Pneumonia	14	10
2. Renal complications	0	5
3. Cardiac complications	2	2
4. Wound complications	0	2
5. Deep metalwork infection	1	0
6. Died in hospital	13	5
Total number of patients with any complication (%)	30 (16.8)	24 (22.2)
Number of patients without complications (%)	149 (83.2)	84 (77.8)
Total number of patients	179	108
Difference in rate of complications (χ^2) p=0.251 (OR=1.42, 95%CI 0.78-2.59)		

Complications were extracted from discharge data. There were more complications in the control group compared to the intervention group. 16.8% of intervention group patients experienced a complication (30 of 179) and 22.2% of control group patients experienced a complication (24 of 108). This was not found to be statistically significant using a χ^2 test (p=0.251). I did not undertake subgroup analysis by complication type here because of the small numbers.

Length of hospital stay

Table 16. Length of inpatient stay by group

	Intervention Group	Control Group
Length of hospital stay (days)		
Mean (SD)	17.7 (14.91)	19.2 (15.89)
Range	2-98	3-113
Difference (t-test) p=0.45 (mean difference -1.455, 95%CI -5.206-2.297)		

The intervention group patients stayed an average of 1.5 days less in hospital than control group patients. This is not found to be statistically significant using a t-test (p=0.45).

Mortality at 12 months

Table 17. 12-month mortality rate by group

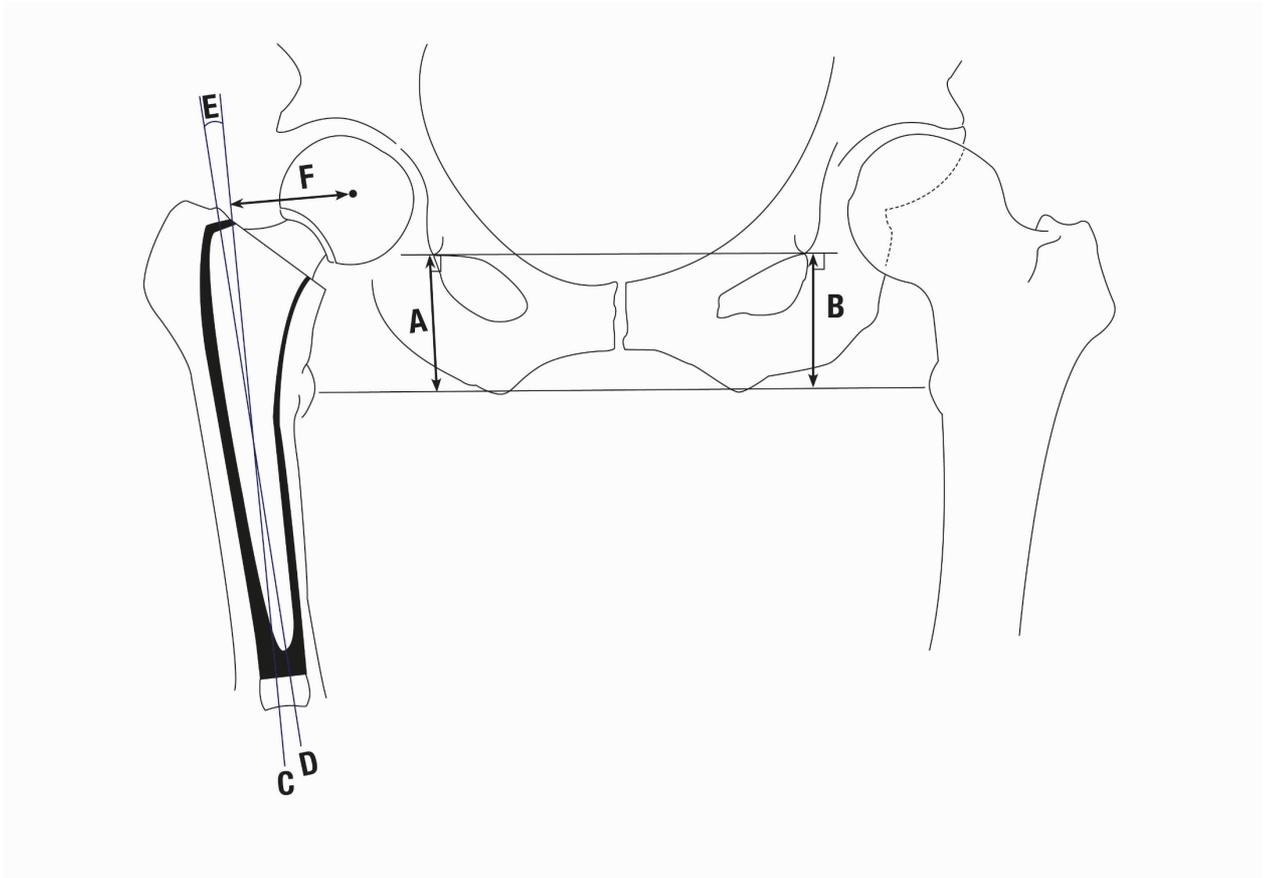
	Intervention Group	Control Group
Died within 12 months of surgery (cases)		
Yes	48	29
No	136	86
Difference (χ^2) p=0.87 (OR 0.96, 95%CI 0.560-1.63)		

There is no significant difference in 12- month mortality rate between the two groups (p=0.87).

2. Hemiarthroplasty

There were 239 hemiarthroplasty operations performed by the study participants during follow up; 160 by the intervention group and 79 by the control group. AP radiographs were available for 151 and 73 cases respectively.

A summary overview of radiographic measurements for hemiarthroplasty is presented below in figure 37.



Leg length discrepancy (LLD) = Distance A (mm) – Distance B (mm)

C = longitudinal axis of femoral shaft

D = longitudinal axis of femoral stem

Angle E = femoral stem alignment (in degrees)

F = femoral offset

Figure 37. Summary diagram of hemiarthroplasty measurements (adapted from the AO Foundation⁴¹²)

Primary Outcome

The primary outcome measure is the quality of operation as measured by leg length discrepancy (LLD).

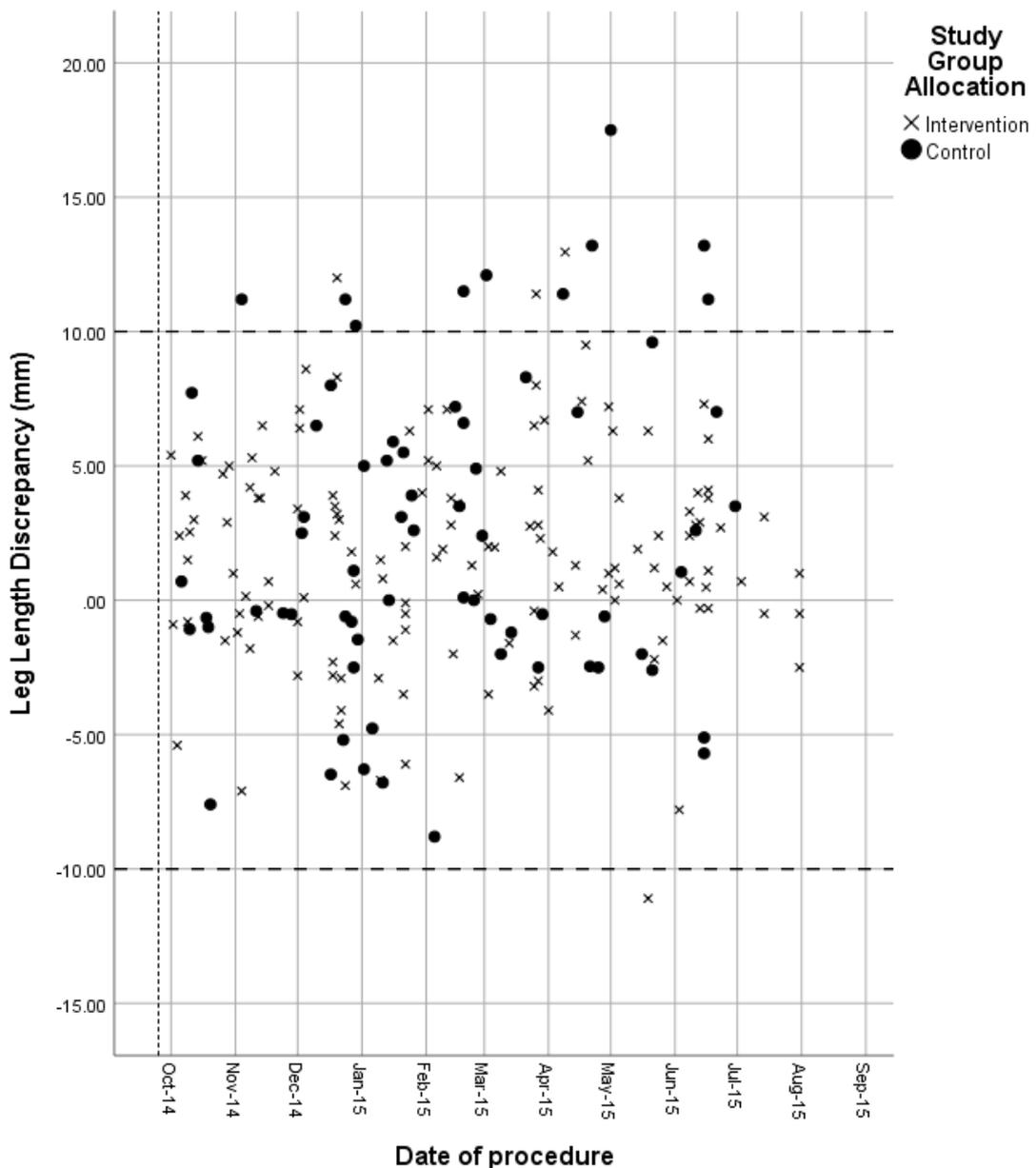


Figure 38. Leg length discrepancy by group

Descriptive analysis shows there were significantly more unacceptable LLD's in the control group compared to the intervention group (2.8% in the intervention group (4 of 142) vs. 14.5% (10 of 69) in the control group, $p=0.001$).

Table 18. Rate of LLD acceptability by group

	Number of patients (%)	
	Intervention Group	Control Group
Leg length discrepancy >10mm	4 (2.8)	10 (14.5)
Leg length discrepancy ≤ 10mm	138 (97.2)	59 (85.5)
Total*	142*	69*
Difference (χ^2) p=0.001 (OR=5.85, 95% CI 1.76-19.39)		

*Nine cases in the intervention group and four cases in the control group had inadequate images to measure LLD.

Secondary Outcomes

i. Radiographic secondary outcome measures

Femoral stem alignment in the AP view

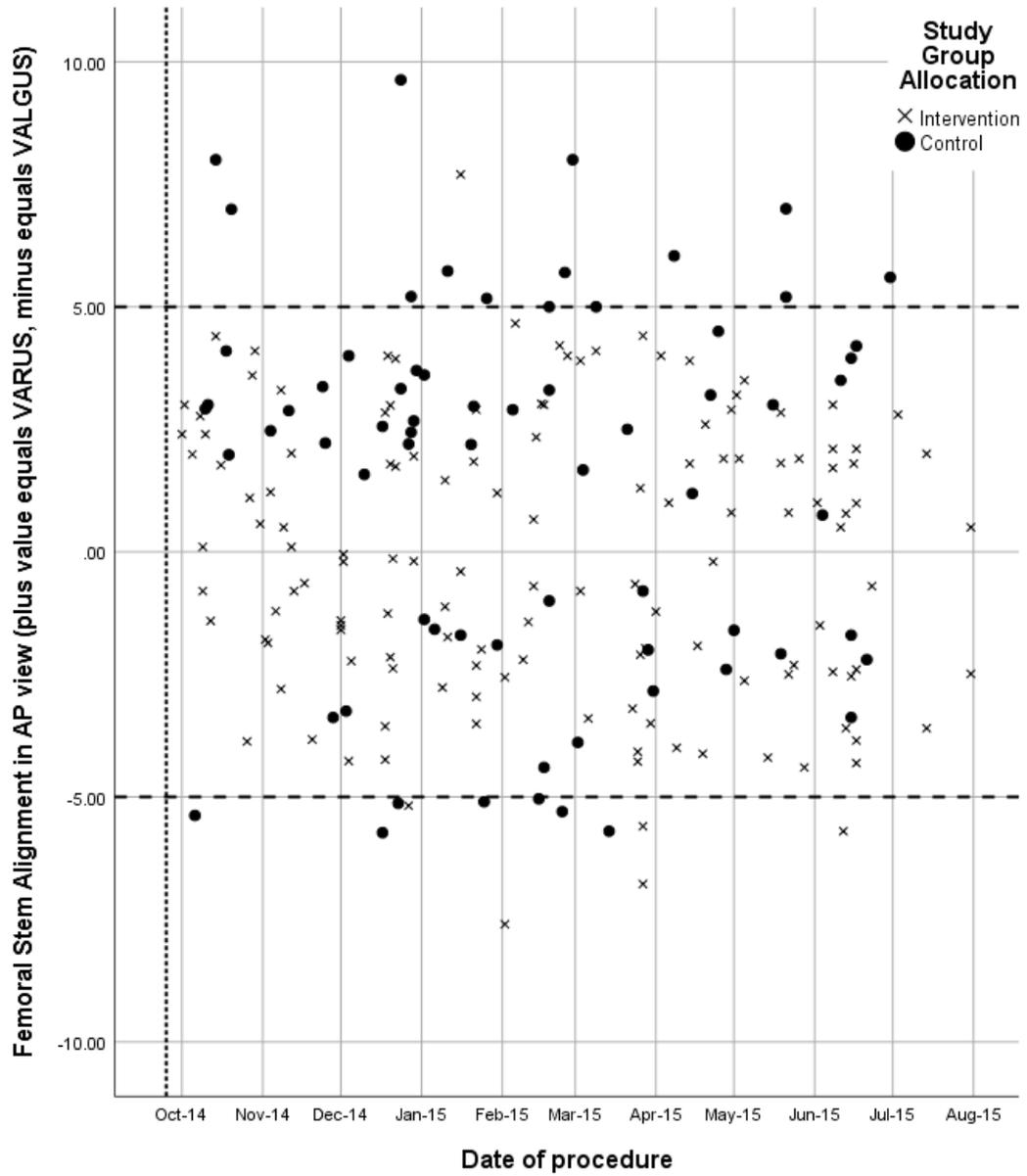


Figure 39. Femoral stem alignment on the post-operative radiograph in the AP view (degrees)

The optimal femoral stem position is in neutral alignment with the longitudinal axis of the shaft of the femur. Varus malposition >5 degrees from neutral in the AP view is associated with failure rates of up to 46% in cemented THA⁴³². It is widely believed that varus malpositioned stems in THA are more likely to dislocate than those in neutral or valgus^{373 463 464}. There are no equivalent implant survival studies in hemiarthroplasty.

Table 19. Number of procedures with acceptable alignment by group

	Number of patients (%)	
	Intervention Group	Control Group
Femoral stem alignment		
>+ 5 degrees from neutral (varus malalignment)	1 (0.7)	14 (20.0)
>- 5 degrees from neutral (valgus malalignment)	5 (3.7)	7 (10.0)
Acceptable alignment (within 5 degrees of neutral)	131 (95.6)	49 (70.0)
Total	137*	70*
Difference (chi-squared) p=<0.001 (OR=9.36, 95%CI 3.56-24.56)		

*Fourteen radiographs in the intervention group and three in the control group were inadequate to measure stem alignment

There are significantly more varus malaligned implants in the control group compared to the intervention group (p<0.001, χ^2 test).

Difference in offset between native and surgical hips

Restoration of offset is important for optimal biomechanical function of the hip and prosthesis survival in total hip arthroplasty^{465 466}. There are no equivalent studies for cemented hemiarthroplasty.

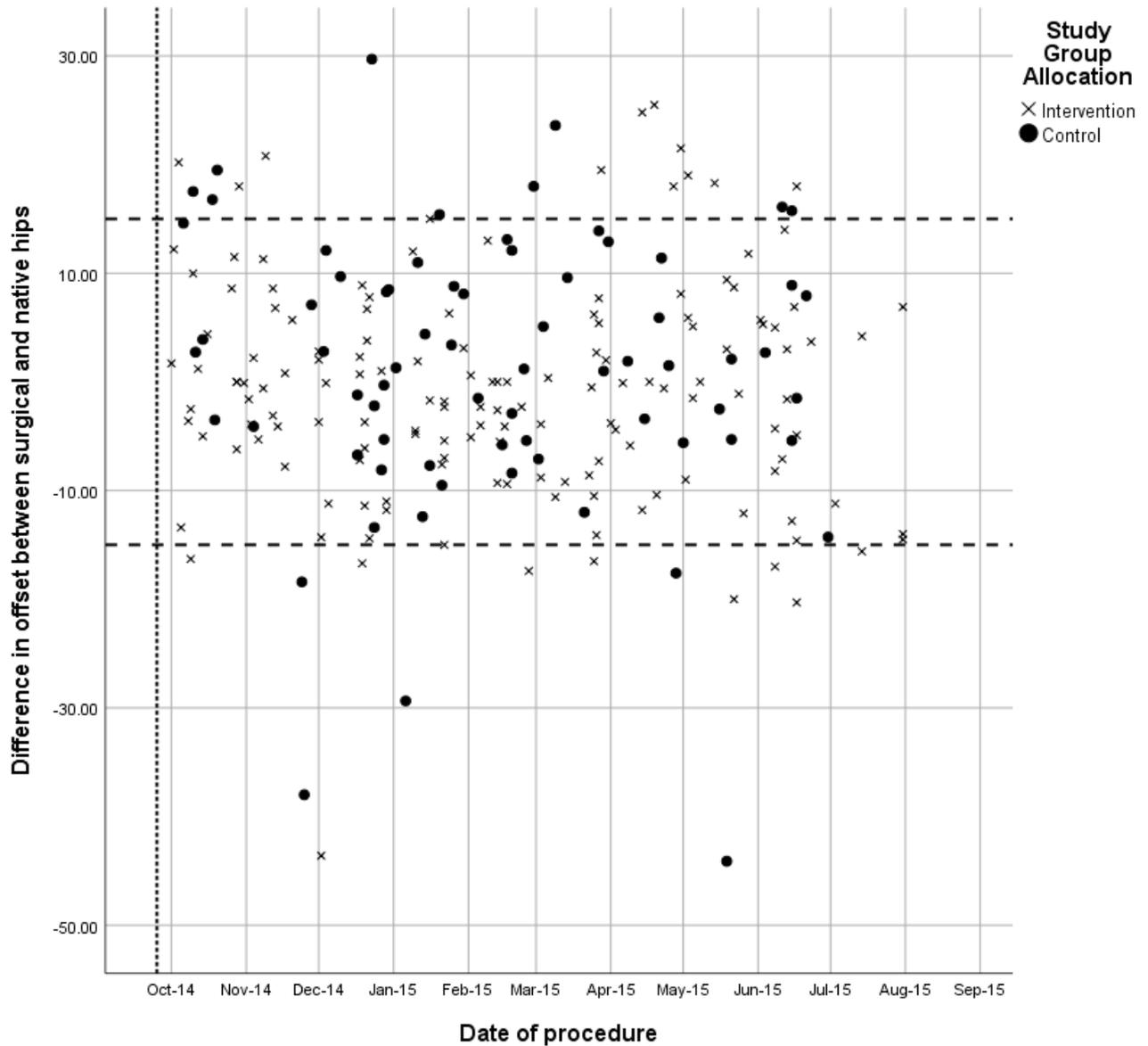


Figure 40. Difference in femoral offset between native and surgical hips

The acceptable difference in offset between native and surgical hips in THA (or hemiarthroplasty) is not currently known. I have arbitrarily chosen to consider a difference of <15mm between the operated and non-operated side to represent a ‘normal’ result, as the vast majority of observations fall within this range.

Table 20. Number of cases with acceptable offset by group

	Number of patients %	
	Intervention Group	Control Group
Femoral offset: number of patients (%)		
≥15 mm compared to contralateral side	22 (14.7)	14 (19.7)
<15 mm compared to contralateral side	128 (85.3)	57 (80.2)
Total	150*	71*
Difference (χ^2) p=0.342 (OR=1.43, 95%CI 0.68-2.99)		

*One radiograph in the intervention group and two in the control group were inadequate to measure offset.

There was a higher rate of unacceptable femoral offset (compared to the contralateral side) in the control group (19.7% of cases) than there was in the intervention group (14.7% of cases). This is not statistically significant ($p=0.342$, χ^2).

Cement Mantle Quality

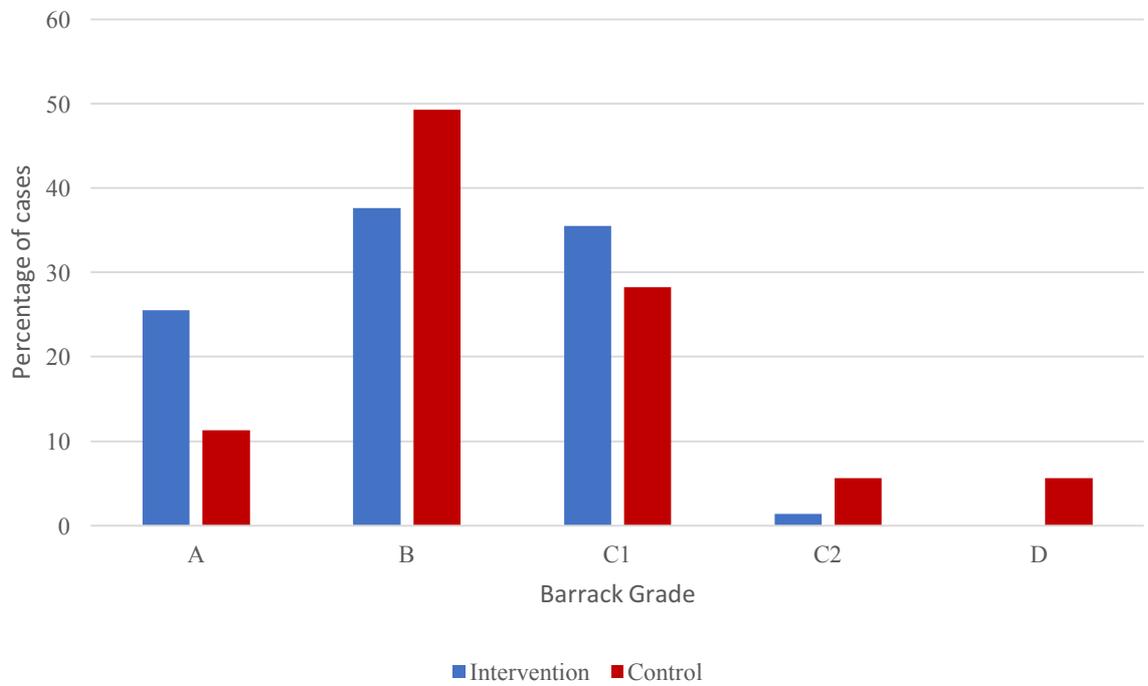


Figure 41. Frequency of Barrack cement mantle quality gradings by study group.

Quality of cementation is a known predictor of implant longevity in THA⁴³³. Poor cementation, as described by Barrack grades C1, C2 or D, is correlated with early implant failure⁴⁶⁷. (Please see radiology manual in appendix 11.5 for detailed description of the Barrack classification).

Table 21. Number of cases of good and poor quality cementation by study group

	Number of cases (%)	
	Intervention Group	Control Group
Good quality cementation (Barrack A or B)	89 (63.1)	41(62.0)
Poor quality cementation (Barrack C1, C2 or D)	52 (36.9)	27 (38.0)
Total	141*	71*
Difference (χ^2) p=0.693 (OR=1.13, 95%CI 0.62-2.04)		

*Ten cases in the intervention group and two cases in the control were uncemented hemiarthroplasties

Rates of poor cementation (C1, C2 and D) were compared between groups. There was a slightly higher rate of poor cementation in the control group (38.0% of cases) as compared to the intervention group (36.9%). This was not statistically significant using a χ^2 test.

ii. Non-radiographic secondary outcome measures

Procedure Time

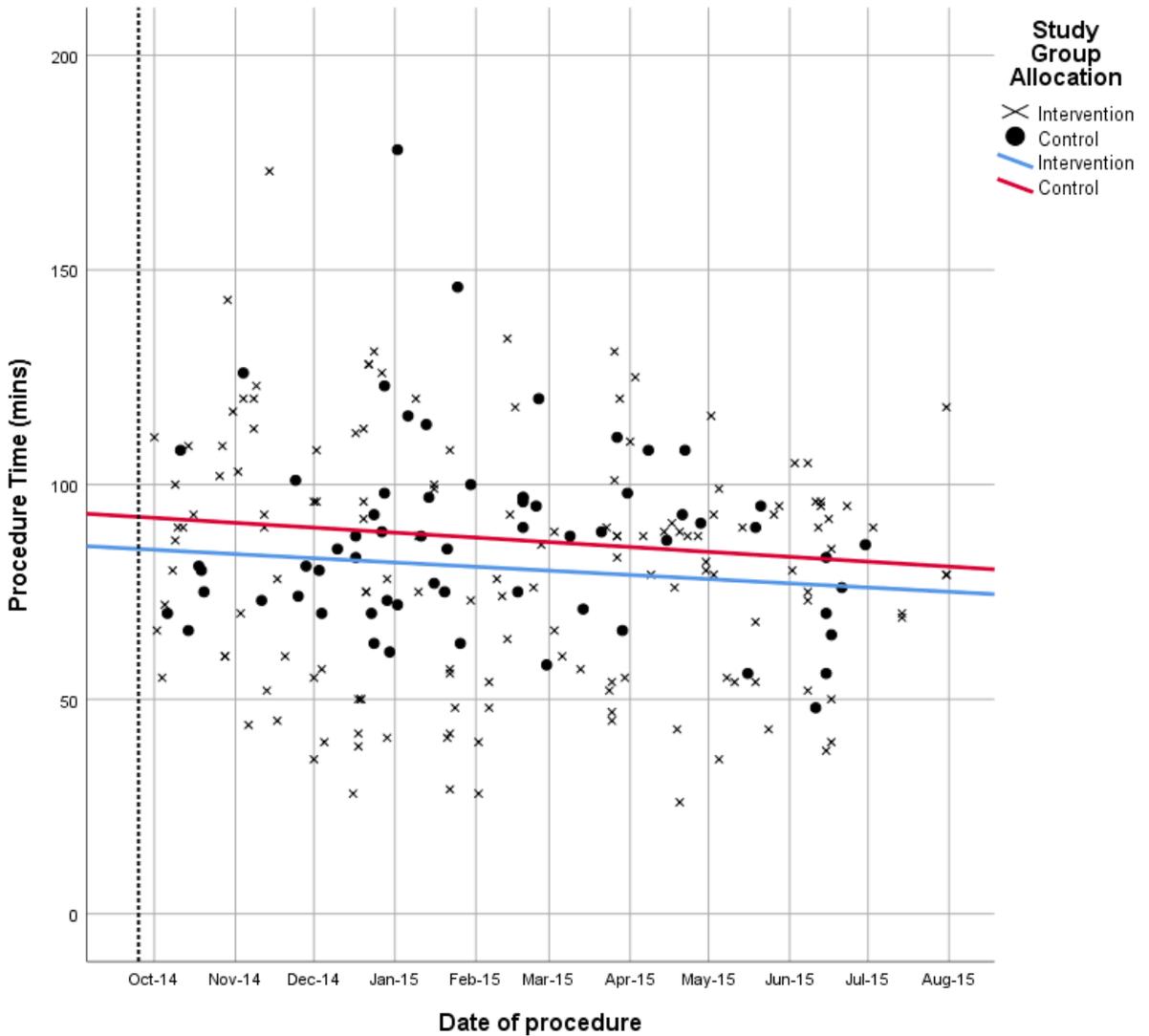


Figure 42. Hemiarthroplasty procedure time by study group allocation.

Mean procedure time in the intervention group was 80.4 minutes (range 26-173 minutes) and was 87.5 minutes (range 48-178) in the control group. The intervention group were on average 7.1 minutes faster than the control group per case. This is not statistically significant (t-test, $p=0.07$, 95% CI -14.77-0.678).

Learning curve analysis shows there was a significant difference between the two groups in the rate of performance improvement as measured by procedure time, i.e.

adding the information about the study arm significantly improved the overall fit of the model ($p=0.04$, likelihood ratio (F-test)).

Intraoperative blood loss

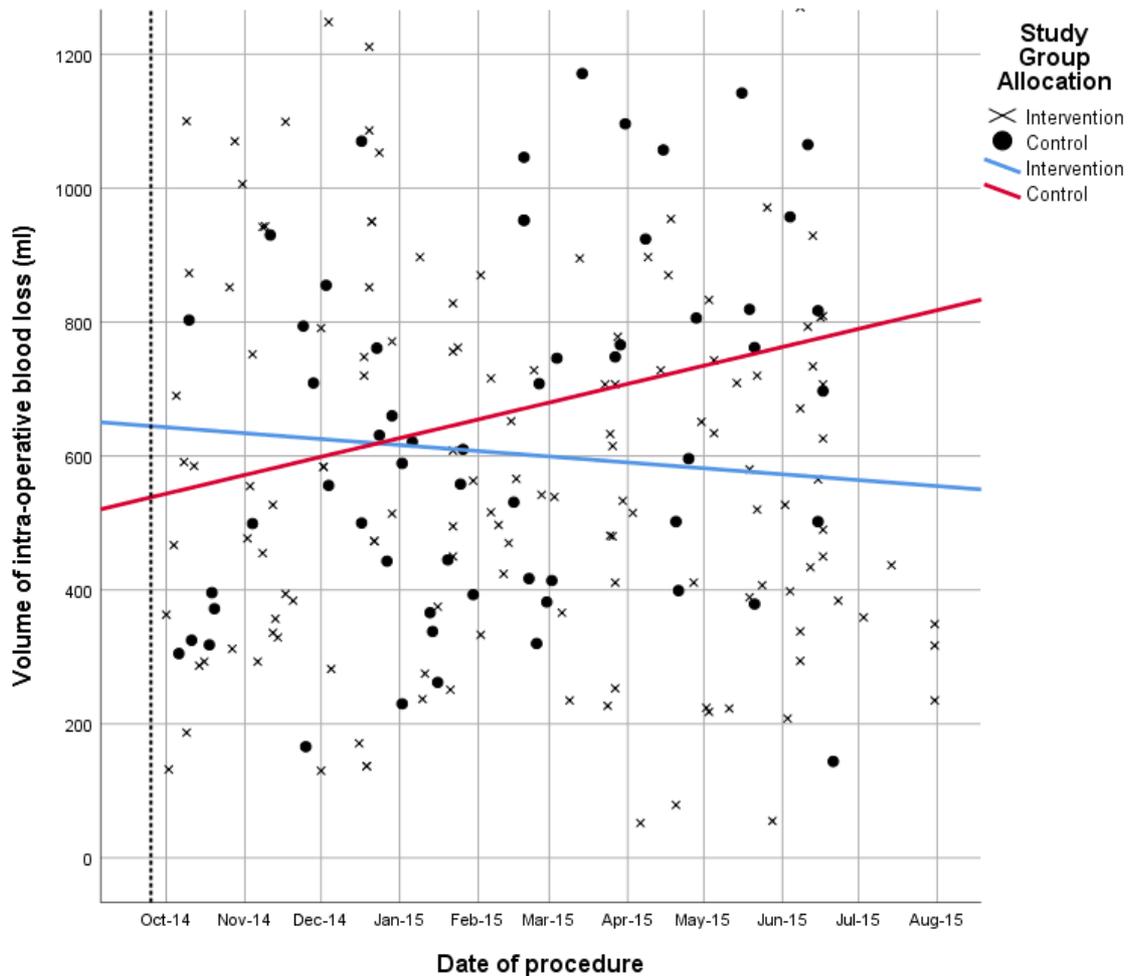


Figure 43. Volume of intra-operative blood loss by group allocation

The mean volume of intra-operative blood loss was higher in the control group than the intervention group: the average blood loss in the intervention group was 603ml (range 52-1248ml) and the average blood loss in the control group was 662ml (range 144-1171ml). This difference was not statistically significant (independent samples t-test, $p=0.21$, mean difference -59.78ml, 95%CI -154.28-34.73).

I compared the rate of blood transfusion within 48 hours of surgery between the two groups.

Table 22. Blood transfusion rates by study group

	Intervention Group	Control Group
Number of patients requiring blood transfusion within 48 hours of surgery (%)		
Yes	6 (3.9)	10 (15.6)
No	146 (96.1)	54 (84.4)
Total*	152	64
Difference (χ^2) p=0.003 (OR=4.51, 95%CI 1.56-13.00)		

*Transfusion data was unavailable for eight cases in the intervention group and 15 cases in the control group

The control group cases were significantly more likely to require a blood transfusion than the intervention group cases (3.9% of intervention cases had a blood transfusion compared to 15.6% of the control group cases, p=0.003, χ^2 test).

Post-operative complication rate

Table 23. Post-operative complication rate by group allocation

	Intervention Group	Control Group
Number of patients with complications;		
1. Hospital Acquired Pneumonia	7	9
2. Renal complications	4	3
3. Cardiac complications	1	0
4. Wound complications	2	2
5. Deep metalwork infection	0	1
6. Died in hospital	4	1
Total number of patients with any complication (%)	18 (12.1)	16(23.2)
Number of patients without complications (%)	131 (87.9)	53(76.8)
Total number of patients*	149	69
Difference in rate of complications (χ^2) p=0.036 (OR=2.19, 95%CI 1.04-4.63)		

*No data for 11 intervention group patients and 10 control group patients.

The control group patients were significantly more likely than the intervention group patients to experience an acute complication during the post-operative admission. The all-complication rate for the intervention group was 12.1% (18 of 149 patients), and for the control group was 23.2% (16 of 72 patients) (p=0.036 χ^2 test). As with DHS, I did not do a subgroup analysis here because of the small numbers.

Length of hospital stay

Table 24. Length of inpatient stay by group

	Intervention Group	Control Group
Length of hospital stay (days)		
Mean (SD)	18.3 (14.92)	22.6(15.30)
Range	2-88	4-86
Difference (t-test) p=0.053 (mean difference -4.30, 95% CI -8.65-0.06)		

The intervention group patients stayed an average of 4.3 days less in hospital than control group patients. This difference borders on statistical significance using a t-test $p=0.053$ (mean difference -4.30, 95% CI -8.65—0.06).

Table 25. 12-month mortality rate by group

	Intervention Group	Control Group
Survived to 12 months post-op: number of patients		
Yes (%)	109(71.2%)	58 (79.5%)
No (%)	44(28.7%)	15 (20.5%)
Total*	153	73
Difference (χ^2) p=0.25 (OR=0.65, 95%CI 0.34-1.25)		

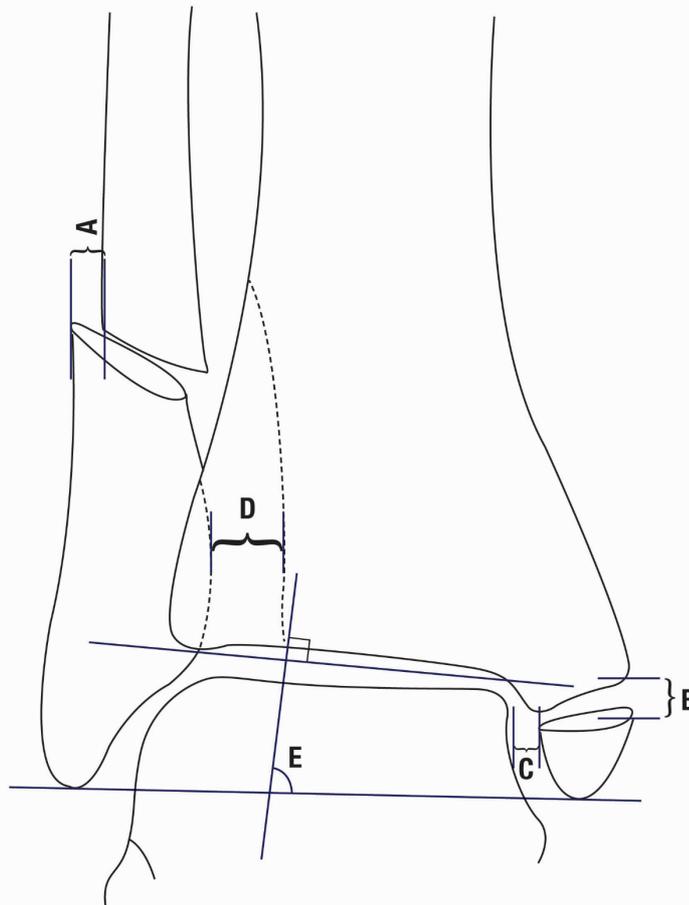
* No mortality data was available for seven intervention group cases and six control group cases.

There is no significant difference in 12- month mortality rate seen between the two groups ($p=0.25$, χ^2)

Ankle fracture fixation

There were 139 ankle fracture fixation operations performed by the study participants during follow up, 91 by the intervention group and 48 by the control group. Adequate post-operative radiographs were available for 81 procedures in the intervention group, and 45 in the control group.

A summary overview of radiographic measurements for ankle fracture fixation is presented below in figure 44.



A = lateral malleolar displacement (mm)

B = medial malleolar displacement (mm)

C = medial clear space (mm)

D = tibiofibular clear space (mm)

E = talocrural angle (degrees)

Figure 44. Schematic overview of ankle fracture radiographic measurements

Primary Outcome

The primary outcome measure is medial clear space.

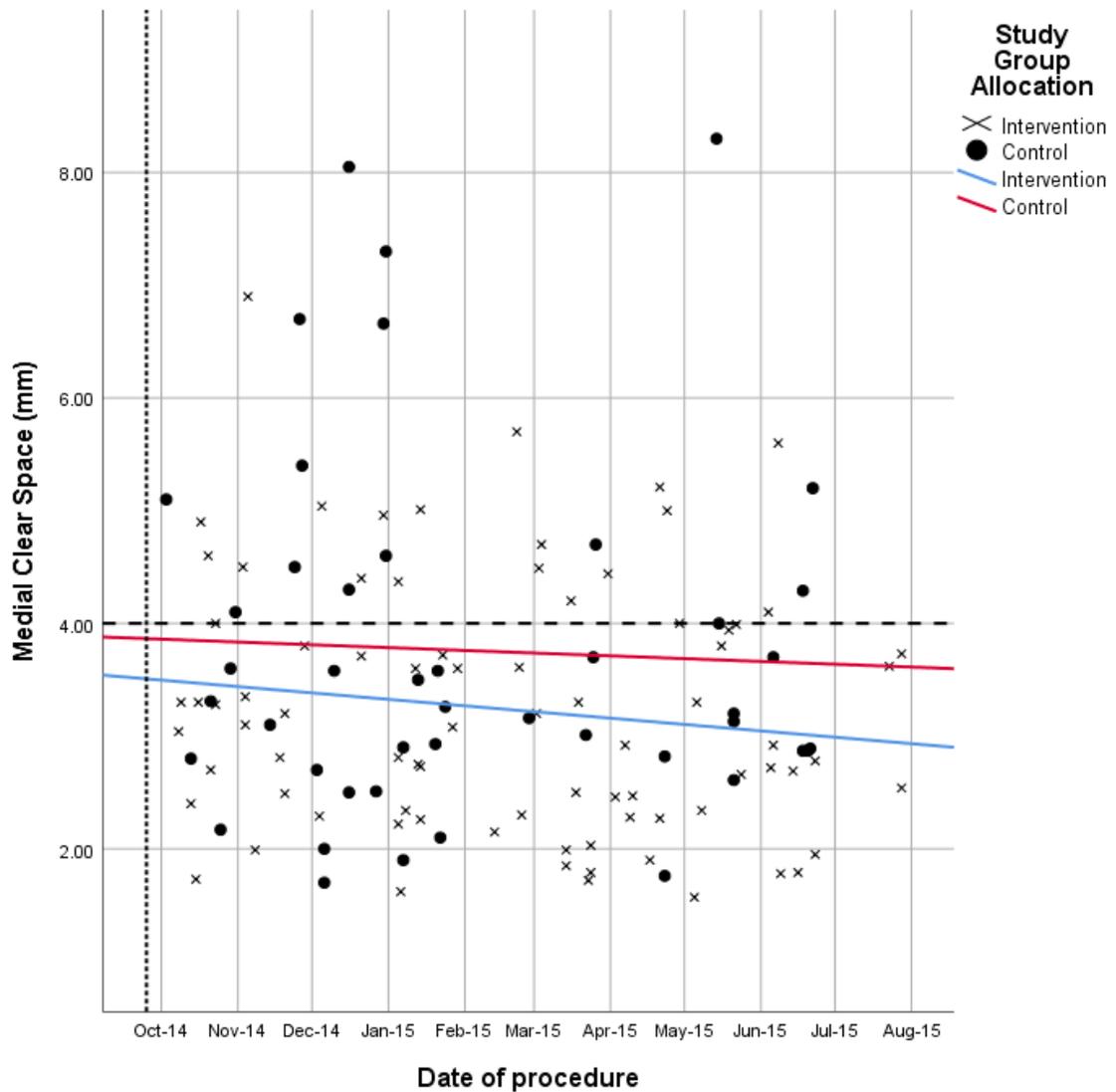


Figure 45. Medial clear space following ankle fracture fixation by group

There is evidence that patients with a medial clear space ≤ 4 mm on the post-operative are significantly more likely to have a good/excellent functional outcome after ankle fracture⁴³⁴. A y-axis reference line at 4mm is shown on the graph (figure 45, above).

Table 26. Acceptability of medial clear space measurements by group

	Number of patients (%)	
	Intervention Group	Control Group
Acceptable MCS $\leq 4\text{mm}$	63 (77.8)	31 (68.9)
Unacceptable MCS $> 4\text{mm}$	18 (22.2)	14 (31.1)
Total	81	45
Difference (χ^2) p=0.272 (OR=1.58, 95%CI =0.70-3.59)		

A higher proportion of control group cases had an unacceptable MCS as compared to the intervention group (22.2% unacceptable in the intervention group (18 of 81 cases), 31.1% unacceptable in the control group (14 of 45 cases)). This difference was not found to be statistically significant using a χ^2 test (p=0.272).

Learning curve analysis

I fitted the linear mixed effects model described above to see if there was a significant difference in the learning curves showing improvement in MCS over time between groups when adjusted for prior experience. This was not found to be statistically significant (p=0.203, likelihood ratio test (F-test)).

Secondary Outcome Measures

i. Radiographic secondary outcome measures

Lateral malleolar displacement

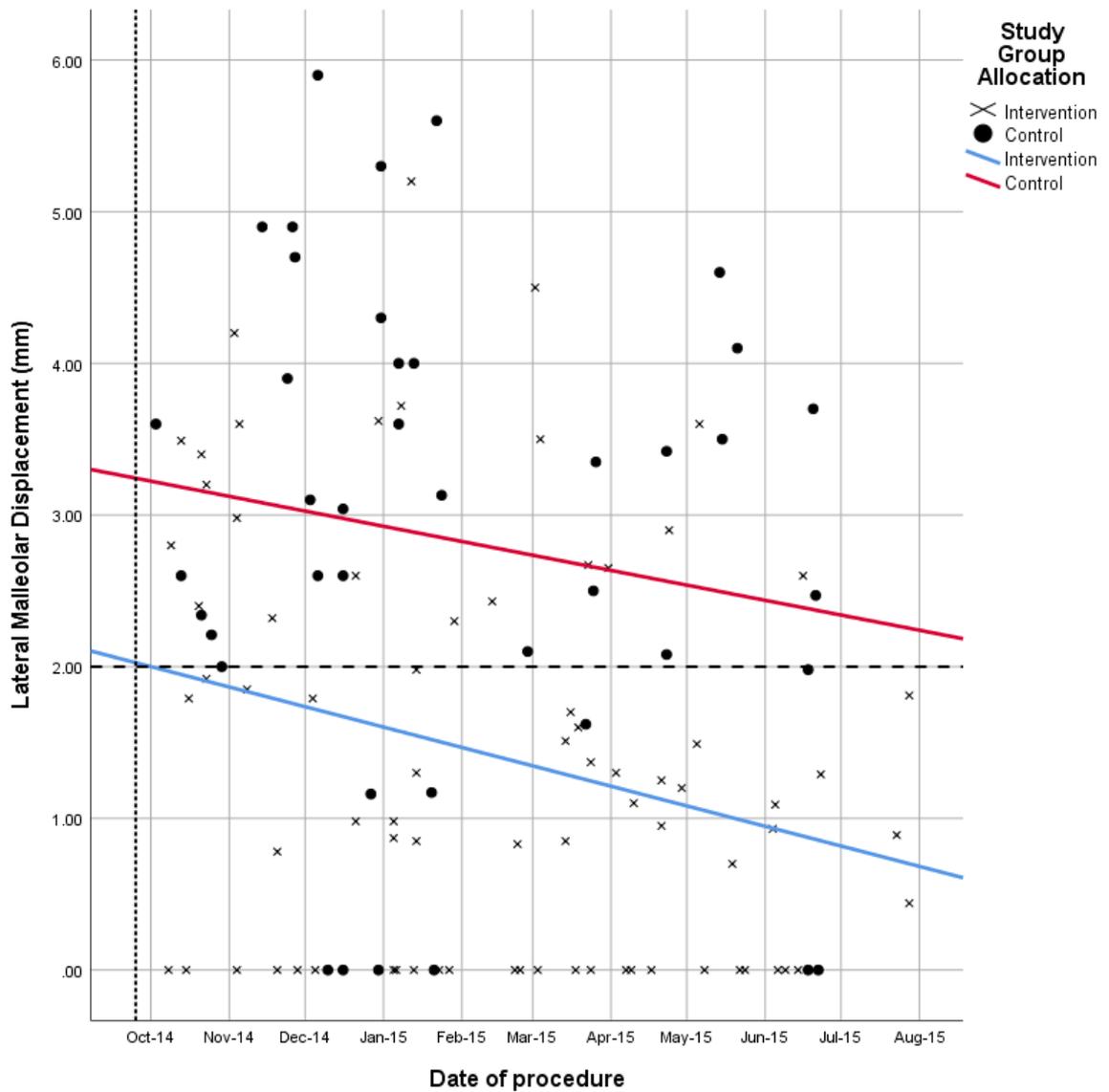


Figure 46. Lateral malleolar displacement over time by study group*

*An LMD of zero equals a perfect reduction of the fracture

Evidence shows that residual lateral malleolar displacement >2mm after surgical fixation of an ankle fracture predicts poor outcome⁴³⁶. A reference line at 2mm is shown on the graph above representing the acceptability threshold (figure 46).

Table 27. Acceptability of lateral malleolus displacement by group

	Number of patients (%)	
	Intervention Group	Control Group
Acceptable LMD ≤ 2mm	55 (71%)	11(27%)
Unacceptable LMD >2mm	22 (29%)	30(73%)
Total*	77	41
Difference (χ^2 test) p=<0.001 (OR=6.82, 95%CI 2.92-15.99)		

*Four cases in the intervention group and four cases in the control group had no lateral malleolar fracture

There were significantly more cases with unacceptable (>2mm) lateral malleolus displacement post-operatively in the control group, compared to the intervention group. 29% (22 of 77 cases) were unacceptable in the intervention group, compared to 73% (30 of 41 cases) in the control group, p=<0.001, χ^2 .

Medial malleolar displacement

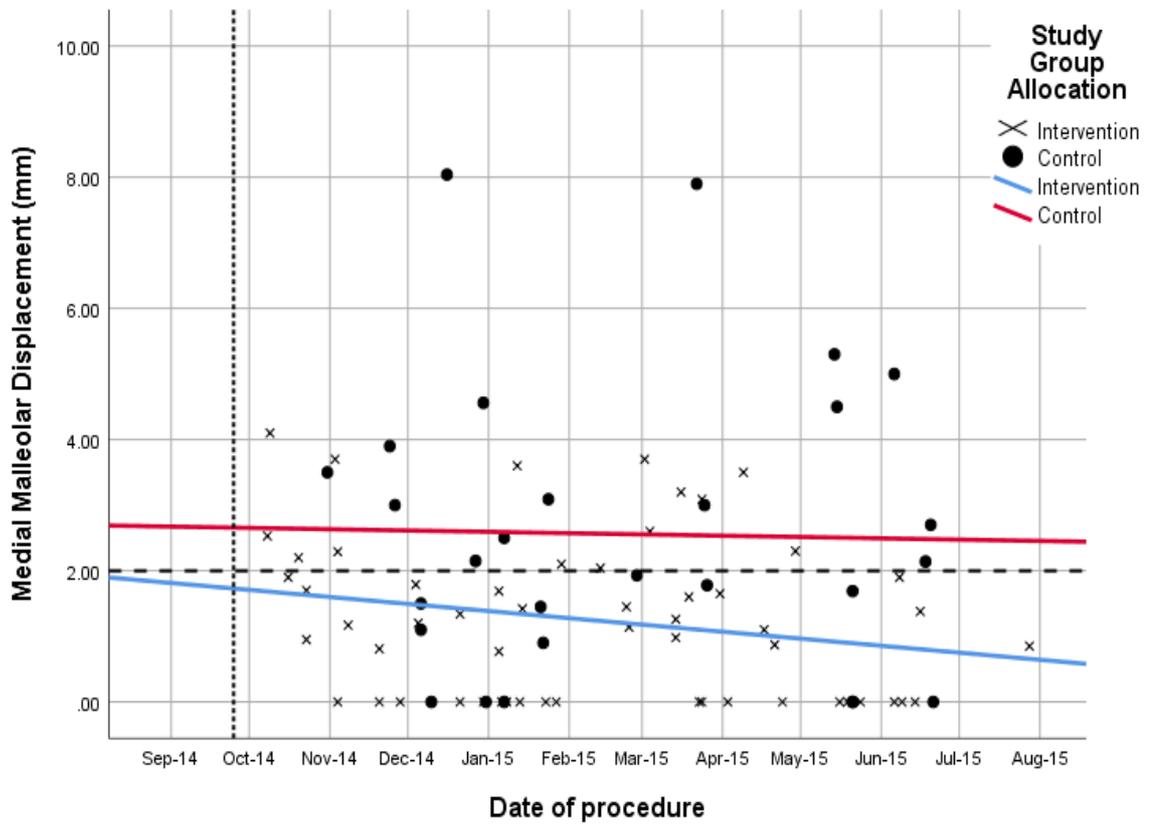


Figure 47. Medial malleolar displacement over time by study group*

*An MMD of zero equals a perfect reduction of the fracture

Evidence shows that residual medial malleolar displacement $>2\text{mm}$ after surgical fixation of an ankle fracture predicts poor clinical outcome⁴³⁶. A y-axis reference line is shown on the graph to reflect this cut-off (figure 47, above).

Table 28. Acceptability of medial malleolus displacement by group

	Number of patients (%)	
	Intervention Group	Control Group
Acceptable MMD \leq 2mm	42 (75%)	13 (46%)
Unacceptable MMD $>$ 2mm	14 (25%)	15 (54%)
Total*	56	28
Difference (χ^2) p=0.009 (OR=3.46, 95%CI 1.33-9.02)		

*25 cases in the intervention group and 17 cases in the control group had no medial malleolar fracture

There were significantly more cases with unacceptable ($>$ 2mm) medial malleolus displacement post-operatively in the control group, compared to the intervention group. 25% were unacceptable in the intervention group, compared to 54% in the control group, $p=0.009$ using a χ^2 test.

Tibiofibular clear space

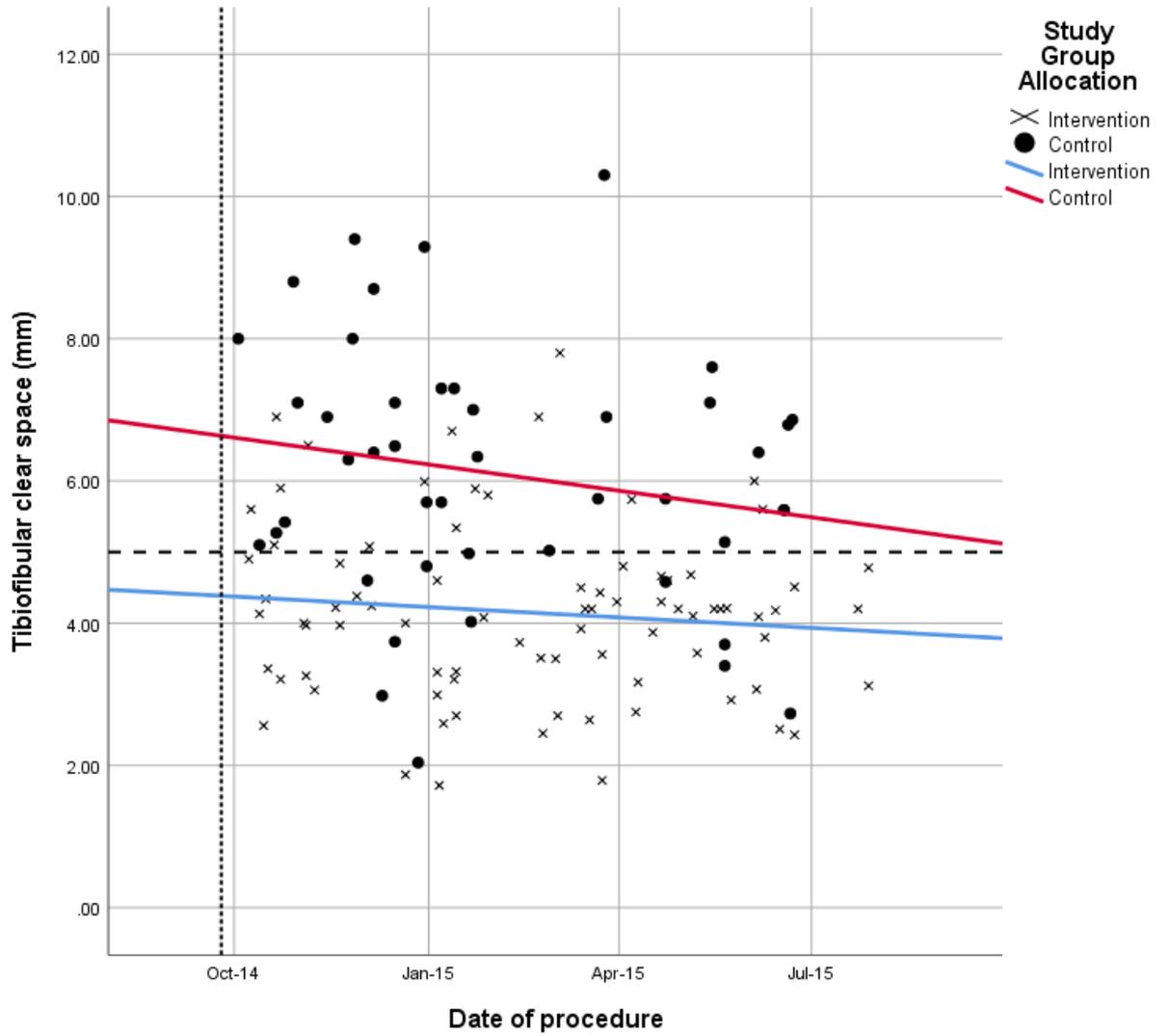


Figure 48. Tibiofibular clear space by study group

The y-axis reference line shows the acceptability cut-off value (5mm).

Table 29. Acceptability of tibiofibular clear space by group

	Number of patients (%)	
	Intervention Group	Control Group
Acceptable TCS (< 5mm)	65 (80%)	11 (24%)
Unacceptable TCS (≥ 5 mm)	16 (20%)	34 (76%)
Total	81	45
Difference (χ^2) p=<0.001 (OR=12.56, 95%CI 5.25-30.05)		

A tibiofibular clear space ≥ 5 mm after surgical fixation of an ankle fracture predicts poor clinical outcome.

There were significantly more cases with unacceptably large (≥ 5 mm) tibiofibular clear space post-operatively in the control group, compared to the intervention group. 20% were unacceptable in the intervention group (16 of 81 cases), compared to 76% in the control group (34 of 45 cases), $p=0.001$ using a χ^2 test.

Talocrural angle

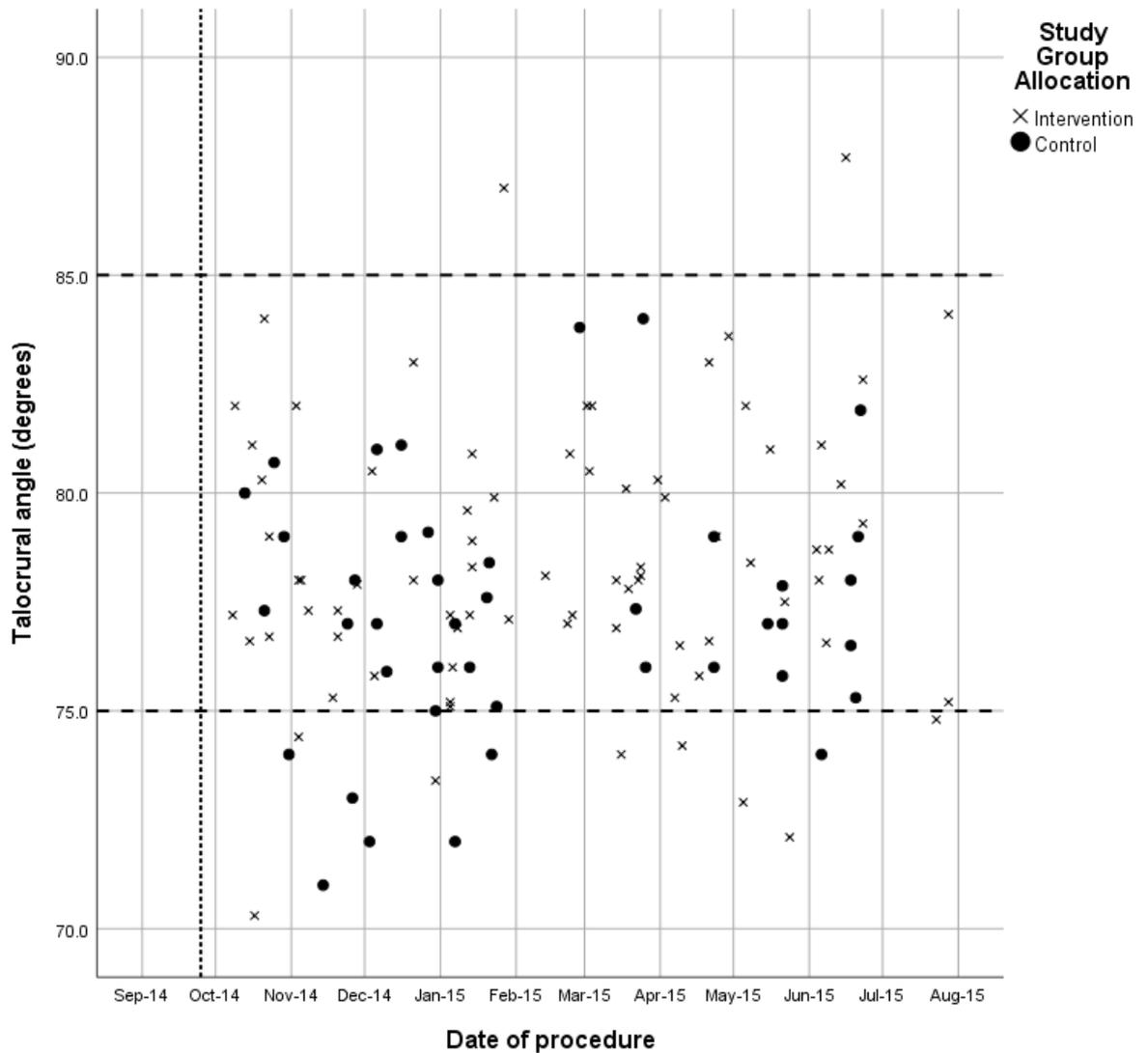


Figure 49. Talocrural angle by study group

Normal talocrural angle ($80^{\circ} \pm 5^{\circ}$) suggests restoration of fibular length after ankle fracture fixation and predicts good outcome⁴³⁷. The acceptability thresholds of $\pm 5^{\circ}$ are shown in figure 49 above as y-axis reference lines.

Table 30. Acceptability of talocrural angle by group

	Number of patients (%)	
	Intervention Group	Control Group
Acceptable TCA ($80^{\circ} \pm \leq 5^{\circ}$)	70 (88%)	35 (83%)
Unacceptable TCA	10 (12%)	7 (17%)
Total*	80	42
Difference (χ^2) p=0.528 (OR 1.40 95%CI 0.49-3.99)		

*Talocrural angle was not measured for one case in the intervention group and three cases in the control group due to poor quality images.

There were slightly more cases with unacceptable talocrural angles in the control group compared to the intervention group (12% in the intervention group (10 of 80 cases), 17% in the control group (7 of 42 cases)), this difference was found not to be statistically significant using a χ^2 test (p=0.528).

ii. Clinical outcome measures

Procedure Time

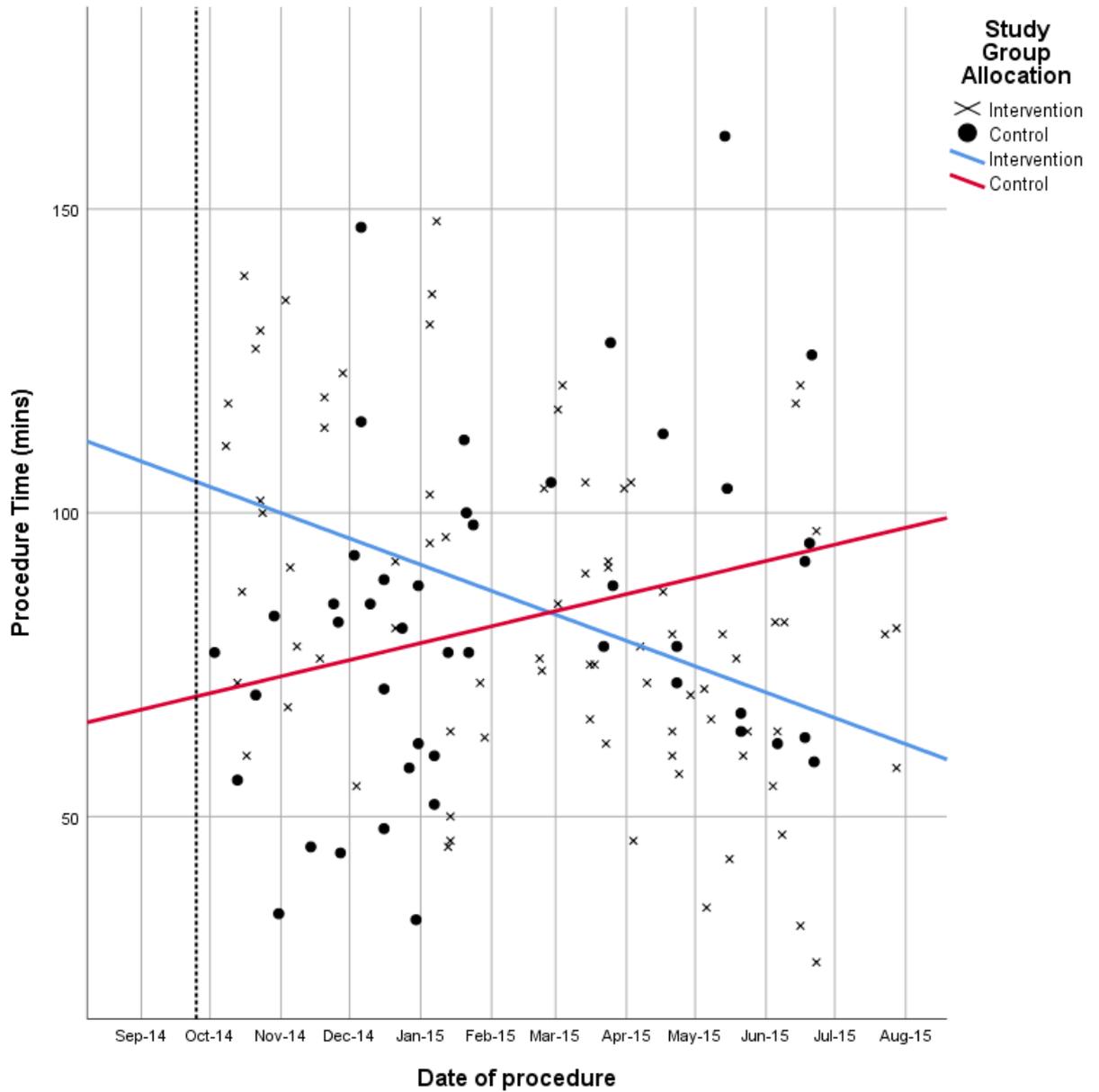


Figure 50. Procedure time by study group

Overall, the control group performed ankle fracture fixation slightly more quickly than the intervention group; intervention group mean procedure time = 83.8 minutes (SD=27.54), range 26-148, control group mean procedure time = 81.7 minutes

(SD=27.70), range 33-162. This difference was not found to be statistically significant ($p=0.66$, 95% CI -8.13-12.33).

Learning curve analysis

I fitted the linear mixed effects model described previously to see if there was a significant difference in the learning curves showing improvement in procedure time over the follow up between groups when adjusted for prior experience. This was not found to be statistically significant (likelihood ratio test (F-test) $p=0.81$).

Intra-operative radiation dose to patient

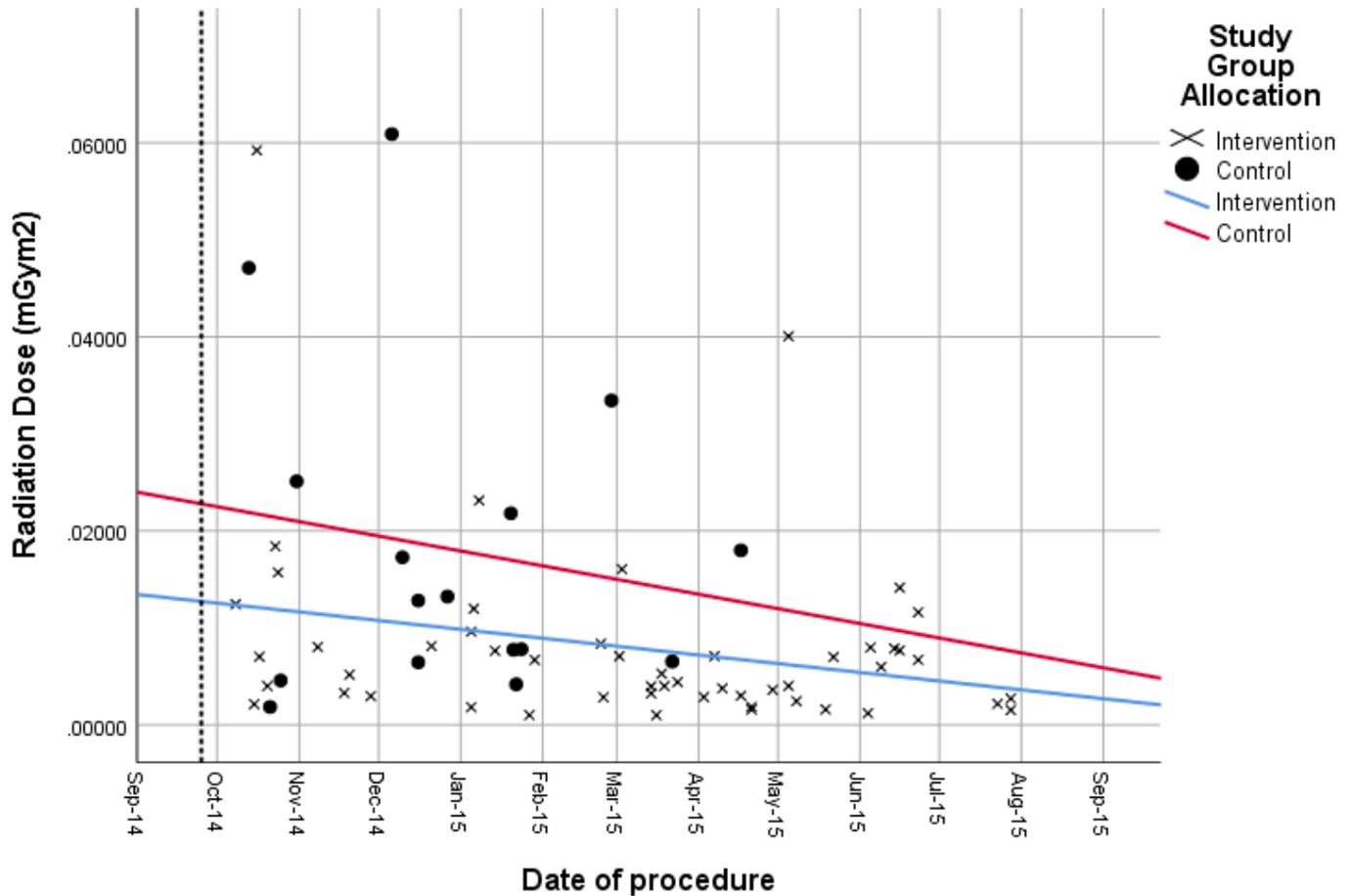


Figure 51. Radiation dose to patient by group over time

The average radiation dose to the patient per case was lower in the intervention group than the control group; intervention group mean dose = 0.0079mGym^2 (SD=0.0099), control group mean dose = 0.0180mGym^2 (SD=0.0166). This was found to be statistically significant using a t-test of independent samples ($p=0.004$, mean difference = -0.0101, 95% CI -0.0169- -0.0034). Learning curve analysis does not show a statistically significant difference in rate of reduction of radiation dose to the patient (likelihood ratio test (F-test) $p=0.301$).

7.5.4 Sensitivity analysis

I performed an intention-to-train analysis (analogous to intention-to-treat in the clinical trials setting) to check the sensitivity of the primary per-protocol ‘as trained’ analysis. The outcomes of this analysis were similar to the results of the per-protocol analysis presented in this chapter (please see appendix 11.7 for summary of results of both analyses and comparisons between the two approaches). I am confident on this basis that the crossovers did not introduce bias.

7.6 Discussion

7.6.1 Cadaveric simulation in T&O improves patient outcomes

The results of this trial show that an intensive CST intervention for junior T&O trainees improves patient outcomes. Cadaveric trained T&O surgeons do better quality operations than standard-trained surgeons, and their patients experience fewer complications.

For DHS, there were significantly fewer malpositioned implants for the cases performed by the CST- trained surgeons, as measured by tip-apex distance >25mm and high +/- anterior screw position ($p < 0.0001$ for both, χ^2). There is a known failure rate for badly positioned DHS implants in the literature (risk ratio 12.7)^{414 458}, and so although I did not collect data on device cut-outs in the study patient population, these can be accurately estimated based on known failure rates.

Device cut-out is a catastrophic complication, requiring re-operation, with significant associated risks to the patient and costs to the health service^{468 469}. Reducing the rate of device failure by cut-out would have substantial economic benefits given the large number of DHS operations carried out annually in the UK⁴⁵³.

Patients whose operations were performed by CST-trained surgeons had fewer acute post-operative complications. This might be because these operations were

performed more expeditiously with a shorter anaesthetic time, which in the frail, elderly population reduces co-morbidity. A better positioned implant, with less soft tissue trauma and muscle bruising from repeated intra-operative instrument readjustments might mean the patient is able to mobilise more quickly post-operatively. Early post-operative mobilisation is known to reduce the risk of acute medical complications associated with prolonged recumbency after surgery, such as hospital acquired pneumonia and thromboembolism⁴⁷⁰.

CST-trained surgeons performed DHS procedures more quickly than standard trained surgeons, and irradiated the patient less. Their patients were also discharged an average of 1.5 days earlier than the standard-trained group's patients. This will also have an associated cost saving to the hospital.

For hemiarthroplasty, the radiographic and clinical outcomes were also better in the CST-trained surgeons' patients.

Leg length discrepancy (LLD) is known to be a predictor of post-operative complications in hip arthroplasty⁴⁶². A lengthened operated leg beyond 10mm compared to the contralateral side can cause pain, premature wear and failure⁴³⁰. There were significantly more unacceptably lengthened legs amongst patients in the standard-trained as compared to the CST trained groups ($p=0.001$).

Femoral stem alignment is also a predictor of implant survival in arthroplasty⁴³². Varus malaligned implants in THA are more likely to dislocate and prematurely fail^{373 463 464}, although there is a complex interplay of patient-related, surgery-related, implant-related and hospital-related factors in determining this risk⁴⁷¹. There were significantly more varus malaligned implants in the standard-trained as compared to the CST trained group ($p<0.0001$).

The CST-trained surgeons were on average 7.1 minutes faster than the standard-trained surgeons when performing hemiarthroplasty. The CST-trained surgeons' patients lost less blood intra-operatively (603ml vs 662ml for the controls), and were significantly less likely to require a blood transfusion within 48 hours of surgery ($p=0.003$).

The acute complication rate was lower for the patients whose operations were performed by CST-trained surgeons, as compared to standard-trained surgeons (acute all-complication rate 12.1% vs 23.2%, $p=0.036$), and they were discharged from hospital an average of 4.3 days sooner.

Superior implant positioning, as measured by leg length discrepancy and femoral stem alignment, will reduce the need for revision surgery due to recurrent dislocation and/or failure. Revision surgery, as with DHS, carries significant risks to the patient and costs to the health service.

The shorter procedure time, lower rate of blood transfusion, lower acute complication rate and shorter inpatient stay seen in the patients of the CST-trained surgeons will all have cost savings to the hospital.

Ankle fracture patients are generally a fitter, younger group. Their journey through surgery is often managed via an ‘ankle pathway’, where a large part of the care is undertaken as an outpatient. For this reason, I did not collect data on length of admission or acute inpatient complications. I also did not collect data on intra-operative blood loss, as the surgery is performed under tourniquet control. Blood loss is therefore usually minimal, with the rare exception of concomitant arterial injury.

It is known that the quality of fracture reduction predicts long term functional outcome⁴³⁴⁻⁴³⁶. A good functional outcome is particularly important in a younger patient cohort who will be placing higher physical demands on their ankle than will the elderly patients on their hip fracture.

There are several radiographic parameters that are known to predict clinical outcome following ankle fracture surgery. These have been previously used in clinical trials⁴²⁸. What is unclear is which of these measurements is most important, or whether a composite score of all measurements is a more valid measure of

outcome^{436 437}. In the absence of convincing evidence for using a composite score, I chose to consider these measurements individually.

Choosing a primary outcome measure for ankle fracture fixation was therefore less straightforward than it was for DHS or hemiarthroplasty, where TAD and LLD were obvious choices because of the body of evidence supporting their validity in predicting outcome after surgery. A large case-series⁴³⁵ in the literature found that if any one of; 1) medial clear space, 2) lateral or 3) medial malleolar displacement or 4) widening of the syndesmosis as measured by tibiofibular clear space were deranged according to the criteria described in appendix 11.5, then an unsatisfactory functional result could be predicted, particularly in younger patients (<40 years). I think it therefore seems reasonable to consider these parameters individually in terms of predicting outcome.

Widened medial clear space (MCS, >4mm) after ankle fracture suggests disruption of the deltoid ligament and ankle instability, which when present after fixation predicts poor clinical outcome⁴³⁵. There were more cases with unacceptable MCS in the standard-trained as compared to the CST-trained groups (31% vs 22% in the CST group) but this difference was not statistically significant.

There were significantly less unacceptable lateral and medial malleolar displacements in the patients of the CST-trained surgeons, compared to those of the standard trained surgeons ($p < 0.001$ for LMD and $p = 0.009$ for MMD, χ^2). The rate of persisting unacceptable lateral malleolar displacement after surgery was 73% in the control group compared to 29% in the intervention group, and 54% and 25% respectively for medial malleolar displacement. This is very high, and suggests that the majority of control group patients are at risk of poor outcomes following surgery according to the Pettrone criteria discussed above.

Tibiofibular clear space (TCS) ≥ 5 mm after surgery suggests persistent disruption of the tibiofibular syndesmosis, and independently predicts poor clinical outcome. There were significantly more cases with unacceptable TSC in the patients of the standard-trained group compared to the CST-trained group (76% vs 20% in the CST

group, $p < 0.001$). In other words more than $\frac{3}{4}$ of control group ankle fracture repairs had evidence of inadequate syndesmosis restoration as measured by TSC. This, in addition to the stark figures for persistent lateral malleolar displacement outlined above, suggests that patients of standard-trained junior surgeons are at a much higher risk of poor outcome following ankle fracture fixation, and that this risk can be reduced using cadaveric simulation training.

Fracture reduction as measured by medial and/or lateral malleolar displacement is known to predict outcome, as malleolar displacement $> 2\text{mm}$ after fracture fixation is associated with increased risk of pain and joint stiffness. Restoration of the talus within the ankle mortise is achieved by accurate reduction of both malleoli (when both are fractured), and they are thought to be equally important in influencing the final results⁴³⁵.

Talocrural angle > 5 degrees from normal (defined as 80°)⁴³⁷ has also been shown to predict poor clinical outcome. Restoration of talocrural angle suggests fibular length restoration and talus reduction within the mortise, so biomechanically this makes sense as a predictor of outcome. There were slightly more cases with unacceptable talocrural angle in the control group than in the intervention group (17% vs. 12% in the standard trained group, but this was not statistically significant).

There was no significant difference seen between the groups in procedure time. The CST-trained surgeons gave their patients a significantly lower dose of intra-operative radiation compared to the controls (0.0079mGym^2 vs 0.018mGym^2 in the standard trained group, $p = 0.004$, t-test). This reduced radiation dose might be more clinically important than in the hip fracture population, as many ankle fracture patients are of reproductive age, and clinicians should be especially mindful of radiation exposure in this group.

7.6.2 Why cadaveric simulation works

Within the caveats of being a small exploratory study (limitations discussed below in section 7.6.5), this trial has shown significant improvements in patient outcome following cadaveric simulation training for all three taught procedures. These improvements are seen in both the technical success of the operation and risk of future failure (as measured by the post-operative radiograph), and in improved clinical outcomes; a reduced acute complication rate (DHS and hemiarthroplasty) and reduced blood transfusion requirement (hemiarthroplasty). There are also some likely financial benefits to the training, seen in quicker procedure time and shorter average hospital stay (DHS and hemiarthroplasty). Further work is needed to undertake a formal health economic analysis to assess the cost effectiveness of the training.

In chapter 2 I discussed the various motor skill learning theories that underpin the surgical learning curve. The cohort of surgeons in the trial were in specialty training years 1-3, and had performed an average of around 12 DHS, 10 hemiarthroplasty and 5 ankle fracture procedures at the start of the study. With the exception of a couple of individuals who were at the novice stage in year 1 and performing these procedures for the first time during the cadaveric training, most participants were at the associative stage of learning (according to the Fitts and Posner model of skill acquisition, chapter 2, figure 1). These individuals were refining established patterns of practice during the course, as advanced beginners (Dreyfus model, chapter 2, figure 2). This intensive deliberate practice ‘in context’, with expert, non-time pressured, bespoke tuition gave the participants the opportunity to progress to a state of competence more quickly than the control group (standard-trained) participants.

An example of how this training translates into patient benefit can be explained by the surgical process in DHS. I have discussed above how TAD predicts outcome, and we can see that there is significant, clinically relevant, superior implant positioning in CST-trained group. But how can CST training lead to an improvement in TAD in patients?

For the benefit of the non-orthopaedic surgeon reader, the position of the screw in the femoral head in DHS procedures (from which the TAD measurement is obtained), is determined by the position of a guidewire that is drilled first into the bone at the early stages of the operation. The position of the wire is then checked in two views using image intensifier. Once satisfactory, a tunnel is reamed over the top of the wire (which remains in place) and the screw is placed over the wire. Only then is the wire removed. Positioning the guidewire is therefore arguably the most critical step of the DHS operation, as it dictates the final screw position (and therefore TAD).

As well as being the most critical step, it is also the most difficult. When a surgical trainee is first learning to do a DHS, it typically takes them several attempts to position the wire correctly. You have to think in three dimensions, do not have direct vision of where the wire is heading in the bone, and are using unfamiliar equipment. It takes a few attempts to realise that dropping your hand actually makes the wire travel upwards, and *vice versa*, and the wire trajectory magnifies any hand movements so any positional adjustments must be subtle. There is a small difference in real life between a good and bad screw position so these micro movements are important. To reposition the wire, it must be reversed back out and a new hole made in the bone. Most junior trainees take several attempts to get the wire position correct, and end up making multiple drill holes in the femoral neck/head in the process, known colloquially as ‘pepper potting’. For obvious reasons, a limit to the number of wire drilling attempts would be set by the supervising consultant in the real-life operating theatre, who might need to take over the procedure, or might sometimes have to accept a suboptimal screw position if the trainee is struggling.

In the cadaveric simulation setting, the trainee can have as many attempts at positioning the wire as he/she needs to obtain the perfect position, and spend a long time doing it if necessary. They can learn about how to make small, repeated positional adjustments, and the supervising trainer can let them make mistakes and start again without the time pressures or safety considerations in the real-life theatre. They could, arguably, learn to drill the wire multiple times on a plastic bone, but there is no soft tissue envelope to deal with, you can cheat using direct vision and

this does not help you much for the challenges of real-world surgery. In the cadaveric laboratory, they can learn that placing the skin incision too far anteriorly will make access to the greater trochanter difficult and they will find themselves struggling with the soft tissues when trying to position the wire drill guide. They can make another, or a much larger incision in this case – which is not possible on the real patient. In the cadaver they can go in and out with the wire as many times as it takes to learn the technique of optimal placement.

This phenomena is probably true for the other taught procedures. Trainees had the time in the cadaveric laboratory, with expert guided tuition, to learn how to correctly position a hemiarthroplasty implant, how to get the version right and how to sink it properly to match the leg length. They could do an excessive soft tissue release so they could see what they were doing properly, and if the implant position was no good then they could do it again and learn from their mistakes. You cannot do this with plastic bones or with computerized virtual reality. You definitely cannot do it on real patients.

The same is probably true for the ankle fractures. By matching faculty sub-specialty interests with the stations (i.e. foot and ankle specialists taught the ankle fracture station), trainers could impart tips and tricks that might elude the non-specialist. The learning opportunity was also effectively doubled by virtue of pairing the trainees, so they learnt from their partner's mistakes as well as their own.

This intensive learning is evidenced by the step-change in ability seen immediately after the training intervention for the CST-trained group for nearly all the outcome measures, with continued improvement seen over the course of the follow-up.

Both groups improved during the follow-up period with real world practice, but the CST-trained group had a head-start because of the training. In other words, the CST-trained participants had been 'upskilled', they had moved further up the surgical learning curve than had the control group, and the fact they were closer to a state of competency is evidenced by the fact that they did better quality operations, had fewer complications, and were generally faster.

7.6.3 Early surgical learning curves can be objectively measured

A secondary objective of this study was to see if early surgical learning curves can be measured objectively.

Several outcome measures showed temporal improvement trends, for both study groups. This shows that there is a learning curve during the surgical training year, and it can be seen in radiographic parameters, procedure time, and image intensifier use. I applied linear sub-group trend lines to the graphs, so the direction of travel of the data can be seen more clearly for the purposes of analysis. Traditionally the learning curve is thought of as literally a curve, and whilst linear trend lines clearly don't allow for this, the temporal improvement can still be seen. I have presented composite learning curves by group, rather than individual surgeon learning curves. This is for the sake of clarity and brevity in this thesis, but an exploration of individual learning curves and the factors that might affect these would be an interesting area for further work.

7.6.4 Post-operative radiographs can be used to demonstrate technical skill

Another secondary objective of this study was to explore the feasibility of using post-operative radiographs to assess the technical quality of surgery using 'final product analysis' (please see chapter 5, table 10 section 4 for supporting evidence).

One of the known challenges in validation work of assessment measures in educational (and indeed clinical) studies is that core instrument properties of validity, reliability and responsiveness are setting and population dependent⁴⁷². Therefore I can say with reasonable confidence that for measuring the technical success of DHS, hemiarthroplasty and ankle fracture fixation operations performed by junior T&O surgeons in the west midlands, the radiographic outcome measures used in this study showed evidence of face validity (they appear suitable for the described purpose, as per the Delphi results), construct validity (they showed a difference between experience level and training type) and responsiveness

(meaningful change over time). I have procedure based assessment scores on some of the cases (not many: approximate mean of 2 per surgeon per procedure type), so there might be a way of measuring concurrent validity, i.e. the extent to which the radiographic measures agree with existing performance metrics. I did not do this as it was not relevant to the aims of the chapter, but is an area for future work.

All the radiographic measurements were taken by me (blinded to group allocation and surgeon identity), so I cannot comment on inter-observer reliability. I only took each set of measurements once (as there were over a thousand cases to measure) so I cannot comment on intra-observer reliability either.

I do have data on the cases coded as ‘assisted’ during the study follow-up, which were effectively performed by consultants, and which do not feature in this analysis for obvious reasons. I could use these in a secondary analysis to benchmark expert performance and examine construct validity in more detail, this is also an area for further work.

The generalisability of these measurements as a core outcomes set for other educational research studies or beyond the research setting for competency assessment in training is unknown. I would like to continue validation work on these as post-doctoral work, as I think they are a promising start towards another objective tool in the technical skills assessment inventory, and may help meet the need for real-world competency based assessment in training (please see chapter 5 conclusion and recommendation section for detailed discussion on this).

7.6.5 Strengths and limitations

As stated above, a considerable strength of this study is that it is the first attempt at providing Kirkpatrick level 4 evidence of the benefit of cadaveric simulation training in open surgery. The randomised design means that unknown confounders should be balanced between the two groups, who were equivalent in the important baseline demographic factors (age, experience level, previous cases).

The multicentre nature of the study design meant that I was able to assess the impact of the intervention in a variety of training environments. About half of the participants were based in one of three regional major teaching hospitals, the rest were working at peripheral district general hospitals. There are advantages and challenges that are unique to the training environments of different types of hospitals (please see chapter 8 for more detailed discussion on this). The different (major trauma centre vs district general hospital) training environments were balanced between the two groups. I can conclude with reasonable confidence that the benefit seen from cadaveric training is not due to some peculiarity relating to one hospital - as could be the case in a single centre study.

I did not do a power calculation, for the reasons as described above, but took a pragmatic approach to sample size. There were several post-randomisation withdrawals in both arms of the study, and a few crossovers, which was disappointing but an inevitable part of real-world educational research. We did not quite achieve our target sample size of 16 surgeons in each arm of the study because of this. There is no evidence to suspect there are any systematic differences between the protocol violators and those who received training as randomized, it was a mere function of pre-existing commitments that could not be rearranged (exams, pre-paid courses etc.).

With hindsight, I think the biggest weakness of this study design was not having a third, low-fidelity comparator arm for the trial. If I were to do a similar trial again I would include this to see if training using low fidelity simulation (e.g. plastic bones), has an impact on patient outcome, and how this compares to standard training and cadaveric simulation.

Caution should also be applied to the interpretation of the meaning of the large observed effect sizes, given the small sample sizes. The fact that the observed (statistically significant) effects are all in the same direction (i.e. favouring the intervention group) and appear stable between a per-protocol and intention-to-treat analysis is however reassuring.

7.6.6 Context of the trial results within the current surgical training climate

We know that there are significant problems with surgical training. As described in detail in chapter 1, these problems can be broadly categorized as quantity problems and quality problems. The quantity problem with surgical training is that due to the change in working hours regulations; there has been an approximately five-fold reduction in the number of hours of training in the space of a generation. A ‘day 1’ consultant in 1990 would have had approximately 30,000 hours of training – compared to just 6,000 hours for a ‘day 1’ consultant in 2020⁴.

The quality problem is that within the limited time there is available in the modern working week, there is tension between the demands of service provision and training. This is exacerbated by the move away from the old style surgical firm to shift-based working patterns, with associated loss of learning opportunities in the elective daytime setting. The increased throughput of complex, multi-morbid patients and loss of foundation doctors mean that junior T&O trainees are spending ever more time on routine administrative and ward medical tasks that offer no training value, at the expense of attending the operating theatre⁹⁴.

Under the current standard practice of ‘master-apprentice’ training, trainee surgeons learn to perform operations on real patients. They will perform their first procedures on real patients as novices, and they will move up the steep part of the learning curve on real patients, to advanced beginner, before reaching competency. The central tenets that are vital for success of the master-apprentice model; large case numbers, long hours of work and consistent mentors are all threatened in the modern training environment.

I believe that the master-apprentice model for the early stage of training surgeons has had its day and is no longer fit for purpose. The early phase of surgical training can, and should, be done in the simulation laboratory, and once competence has been achieved then learning can be refined during procedures on real patients.

7.7 Conclusions

Hip fracture is a major public health burden. About 65,000 hip fractures occur annually in the UK⁴⁷³, 98% of which require surgery⁴⁵³. The cost to the NHS for managing hip fractures (not including social care) is £1 billion⁴⁷³. Most of these operations will be performed by junior T&O surgeons in training.

There were significant improvements in implant position (with associated reduction in risk of failure) and a reduced acute complication rate seen in the hip fracture patients whose surgeon had undergone cadaveric simulation training. They also had a shorter operation time and shorter inpatient stay after surgery. They lost less blood intra-operatively and were less likely to need a blood transfusion (hemiarthroplasty). Given the scale of the problem and cost of hip fracture in the UK, even a small reduction in risk of device failure or acute complications is important and relevant.

The trial results suggest there could be substantial NHS cost savings to be had if all T&O junior surgical trainees in the UK underwent CST training. Formal health economic analysis is needed to put some numbers to this assumption.

Nationally co-ordinated cadaveric simulation training programmes are already underway in Neurosurgery and Vascular surgery. I can see no reason why Trauma & Orthopaedics cannot do the same – **no trainee surgeon should be doing an operation for the first time on a living patient.**

7.8 Reflections

7.8.1 Trial results

Re-reading this chapter leaves me with mixed feelings. On the one hand, I am delighted that the training appears to have worked so well, and has benefited the patients. This is really exciting, and as the first Kirkpatrick level 4 study showing patient benefit from cadaveric simulation training for open surgery I hope these results will be of wide interest to the surgical education community. I also hope that it will contribute towards the necessary high quality evidence to justify expanding CST provision in T&O.

On the other hand, I was also surprised and disappointed to see that there are still obviously substandard operations being performed on patients by junior surgeons, for lack of proper training. Patients are actually suffering the consequences of the problems that have been extensively described in a hypothetical sense with surgical training.

I found some of the numbers quite shocking in this regard. Almost $\frac{3}{4}$ of ankle fracture fixations performed by the control group had at least one 'inadequate' reduction parameter (as per Pettrone criteria), risking poor functional outcome. One in five control group DHS patients had a TAD that puts them at more than twelve times the baseline risk of screw cut-out and re-operation. One in seven control group hemiarthroplasties had an unacceptable leg length discrepancy, putting them at greatly increased risk of pain, dislocation and failure.

There is also an interesting seasonal effect whereby all outcomes seem to improve slightly later in the training year, regardless of group allocation. This shows that the known 'Black Wednesday' effect in medicine (where medical inpatient mortality and complication rates increase in August due to junior doctor change over) also appears to apply to the quality of surgery. To my best knowledge, this has not been reported before. I didn't go into a formal analysis of this in the chapter as I thought

it was not relevant to the aim of the trial. There is more secondary analysis to be done here in the future.

7.8.2 Running a trial

Before starting this project, I had underestimated the amount of work involved in running a trial. Having now occupied almost every role on a typical trial team in doing this (small) study, I have a substantial appreciation and respect for why the large clinical trials require so much manpower and financial resource.

Organising the CST course was a considerable task, and I am grateful to my supervisors for their support with this, and to our consultant faculty who gave up their professional development time and/or days off to teach on the course. The original fellowship award covered my salary, but there was no provision for consumables. I applied for a project grant from the deanery, and also an independent charity. Fortunately both applications were successful and we had about £75,000 available to spend on the two courses. Thanks to industry connections from my supervisors who donated implants, drapes and instruments, and a generous in-house discount offered by the surgical training centre, we were able to come in well under budget and only used about half of the money on the two courses. Our biggest cost by far was the cadaveric material which was imported under license from the USA.

The delivery of the CST course – a critical time for the trial – fell in September 2014 at the same time as my first daughter was born. I was back to organising the trial 2 days after she arrived, and then overseeing the running of the course in the surgical training centre and doing the consent and study briefings two weeks later. I am not sure with hindsight that this was very sensible (I had somewhat underestimated the impact of having a new born baby), but I am one for applying maximum effort to my endeavours. In the end the course went smoothly and the trial launched successfully.

Collecting the data was especially challenging. It took over two years to get the project through ethics approval, in a large part due to extraordinary delays (months at every stage) in processing the application by the Confidentiality Advisory Group.

Fortunately for me, part of this delay was absorbed as I was by this point on a years' maternity leave with my second daughter, but had I been doing this PhD on a traditional 3-year full-time trajectory then we would not have got the approval in time to collect the data. It was at times enormously frustrating.

Another significant delay, once we finally had ethical and CAG approval, came with requesting the electronic logbook data from eLogbook. This data was obviously critical as without it I could not identify the cases performed by the participating surgeons.

The eLogbook data manager, who handles extraction requests, left their post suddenly without warning once we were in the advanced stages of the request (final amendments to an approved-in-principle data application). What followed was months of frustration wherein the post remained vacant, eLogbook staff would not entertain or even discuss the request, and the months ticked by with no progress. Some of this delay was fortunately absorbed by yet more maternity leave (my third daughter in three years). I was able to keep busy with writing the other chapters in this thesis, but there was a real low point around the autumn of 2018 when we didn't think the data was going to be obtainable, the project might fail, and some sleepless nights ensued. Fortunately, eventually, I finally managed to get the data.

Once we knew where and when the operations had taken place, the next challenge came in accessing the radiographs and clinical data from the nine study sites. I had assumed that once we had ethics approval, and had the case information from eLogbook, that accessing the hospital-based data would be fairly straightforward. However it was, in fact, a bureaucratic nightmare.

I applied through the respective Research & Development departments of the hospitals, each of which had their own sets of lengthy, complicated and slightly different versions of forms to complete. Some sites were very responsive, others were not. The slowest site took more than 5 months to reply to my initial contact, despite regular chasing.

There had also been some restructuring of the Health Research Authority during the time it had taken to get to this point, and so I had to complete a new Health Research Authority (HRA) registration application and new set of event schedules and statement of activities, approval for which took a further 8 weeks. Once each of the sites had eventually approved the request and confirmed capacity and capability, I then had to find a host clinician at each site to act as local project PI. This was straightforward in some sites and much more difficult in others. We could not have recruited local PI's at the planning stage as we did not know which hospitals our participants would be working at until we had access to the eLogbook data.

It took months of legwork, hundreds of emails and phone calls, and many files worth of paperwork to finally access the data. I found being dependent on what was clearly a very inefficient and 'one size fits all' research governance system for the success of the study to be stressful and frustrating. The day the last bit of data came in and the trial database was finally complete, more than five years after starting the study, was a relief to say the least. The governance arrangements felt especially heavy handed given that this was a low risk educational, and not clinical, trial. A more rational research governance system would have in-built flexibility to stratify applications by risk, and allow for a more streamlined 'light-touch' approval for low risk projects. Standardizing the processing of these requests across different NHS sites would also help to improve efficiency.

On a personal level, I feel that my main achievement with this project has been successfully navigating through the seemingly endless and paralyzing research bureaucracy. It has without doubt been a useful training (and resilience) exercise, however it is not an experience I ever wish to repeat. Any future trials of mine will be done with the proper support of a professional trials team!

Chapter 8: Cadaveric simulation for orthopaedic surgeons-in-training: how does it influence learning and why? A qualitative study

In this chapter, I will describe the qualitative component of my research into the impact of CST on postgraduate orthopaedic surgeons-in-training. I felt a qualitative enquiry was necessary for an in-depth exploration of the experiences of the surgeons undertaking the training, their perspectives of its value (or not) as a training tool, and to begin to develop a more sophisticated understanding of how skills learnt in a simulated environment are translated into the workplace from the perspective of the lived-experience of the participants. I also wanted to explore some of the wider issues around the role of simulation in surgical training, to develop a richer understanding of the context in which the quantitative aspect of this project sits.

Declarations

This chapter has been published;

James HK, Pattison GTR, Griffin DR, Fisher JD. How does cadaveric simulation influence learning in Orthopedic residents? J Surg Educ. 2020;77(3)671-682.

James HK, Fisher JG, Griffin DR, Pattison GTR. If we can't get to theatre, we can't learn to operate. A qualitative study of factors influencing core trainee access to the operating theatre in Trauma & Orthopaedics. Ann R Coll Surg Engl. 2020, In Press.

8.1 Introduction

The literature on the qualitative impact of simulation training for surgeons is limited. As demonstrated by the literature review presented in chapter 3, surgical simulation research is generally undertaken from a ‘quantitative, reductionist perspective’⁴⁷⁴. This is probably because the measures of ‘effectiveness and efficacy of training’ relate primarily to the mastery of technical skills and confidence gains, including ‘downstream translational effects on patient outcomes and patient care practices’⁴⁷⁴.

In terms of providing objective evidence of impact, this approach makes sense. From a broader, and more holistic viewpoint however, a criticism of restricting simulation-based research to purely outcome and effectiveness-based studies is that our understanding of the benefits of simulation-based education remains mono-dimensional. Qualitative evidence can offer richer and more contextualised detail of impact, and provide explanation of phenomena that could be used to improve the educational value of future simulation training activity.

This qualitative enquiry is the first study to begin to formally explore the impact of cadaveric simulation training on postgraduate surgical trainees. The purpose of this chapter is to expand on and provide context to the findings presented in chapter 7, by exploring the participants’ experiences of the CST intervention.

8.2 Research Questions

- 1) How do postgraduate orthopaedic surgeons-in-training experience an intensive cadaveric simulation training course?
- 2) How are skills that are learnt in a simulated environment translated into the workplace?
- 3) What are the factors affecting access to traditional-style training in the operating theatre?

8.3 Paradigms: Ontological and Epistemological perspectives

I have approached this research from an experiential perspective, as I am seeking to validate the meanings, views and perspectives of the participants, as expressed in the data⁴⁷⁵, and the participants interpretations of their experiences are accepted and prioritised in the analysis. A wholly qualitative approach enabled me to focus on the participants own framing around issues, and their own terms of reference, allowing a fuller, multi-faceted understanding of the issues around the use of CST, in an exploratory and flexible manner that embraces the complexity of human experiences and perspectives.

A declaration of my ontological and epistemological assumptions as researcher in conducting this inquiry is necessary to justify my approach to answering the research questions.

Ontology refers to the ‘nature of being’⁴⁷⁶, and to specify ‘the relationship between the world and our human interpretations and practices’⁴⁷⁵. A recognized continuum exists between, at one end, ‘realism’, where there is one absolute truth that is entirely independent from human ways of knowing and interpreting it, which can be accessed by the application of the correct research techniques. At the other end of the ontological spectrum is ‘relativism’, where reality is ‘entirely dependent on human interpretation and knowledge’⁴⁷⁵. A relativist ontological position argues there are multiple constructed realities, which are time and context dependent and which can never be fully accessed or known (Cromby & Nightingale 1999 as cited by Braun & Clarke⁴⁷⁵). Typically, a realist ontological position of ‘one absolute knowable truth’ underpins most quantitative medical research, and is the position with which I am most familiar in my background as a surgeon trained in principally quantitative research methods during my undergraduate education at medical school. A middle ground between the realist and relativist positions exists, which is the critical realist position.

Critical realism accepts there is an absolute reality and ‘pre-social’ knowable world, but this sits obscured behind the ‘subjective and socially-located knowledge’ that is accessible to the researcher (Madill 2000 as cited by Braun & Clarke⁴⁷⁵), hence we

can only ever partially access reality through our research efforts. I have taken a critical realist position in the conduct of this project; my ontological assumption here is that there is a true reality of how CST delivery impacts surgeons-in-training, and how skills taught in simulation are transferred into the workplace, but this is partially discoverable as I can only access this truth through the lens of the participants' own lived-experiences and perspectives, and my lens as researcher.

Epistemology refers to 'theory of knowledge'⁴⁷⁶, and addresses the issue of 'what it is possible to know'⁴⁷⁵. At one end of the spectrum, 'positivism' assumes 'a straightforward relationship between the world and our perception of it'⁴⁷⁵, and that the researcher can 'empirically measure reality'⁴⁷⁶. This position dominates the field of quantitative medical research, and is the epistemological position doctors are trained in from an early career stage, with the belief that the truth is absolute, and discoverable with the right research procedures.

In contrast, 'constructionism' holds the opposite view that knowledge is constructed out of the social world we live in, and hence seen as 'social artefact', which is relative and influenced by social, moral, political and ideological (etc.) variables. Contextualism is an epistemology that sits between positivism and constructionism in its assumptions. This contextualist position centres around the belief that knowledge derives from contexts and reflects the researcher's position, i.e. is 'local, situated and therefore always provisional' (Madill et al 2000 and Tebes 2005 cited by Braun & Clarke⁴⁷⁵) but retains a determination to understand truth, and hence has a 'realist dimension'⁴⁷⁵. Contextualism asserts that there is no single research method that can access the truth (Tebes 2005 cited by Braun & Clarke⁴⁷⁵), however knowledge is valid in certain defined contexts.

Arguably the most common epistemological position within medical education research, and that with which I am most familiar, is post-positivism. Post-positivism shares with positivism the 'belief that there is an objective reality that can be discovered', but is distinguished from pure positivism by the acknowledgment that 'complex human behaviour is shaped by individual motivations and cultural environments' and that research effort must 'represent these complexities rather than elide them in the search of a contextual essence or truth'⁴⁷⁶. I have taken a mainly

post-positivist position in the conduct and analysis of the quantitative part of the research, and a more contextualist stance for this qualitative enquiry.

8.4 Methodology

8.4.1 System of enquiry

I have used experiential thematic analysis in this research. I chose the experiential variety of thematic analysis because the aim of the research was to focus on the participants' experiences and perspectives of CST.

8.4.2 Method of data collection

I conducted semi-structured interviews with the intervention group surgeons (n=14) who were enrolled in the CAD:TRAUMA randomised controlled trial (please see chapter 7). I developed and piloted a topic guide to help provide a degree of structure to the interviews, to help me as a novice interviewer to keep the interviews broadly 'on topic' (appendix 11.8).

Individual interviews were chosen over a focus group for both academic and pragmatic reasons. I felt that from a research point of view, interviews would yield rich and detailed data about individual experiences, offering the flexibility to expand on interesting or unexpected issues raised and explore any sensitivities that may arise. From a pragmatic point of view, the study participants were all full-time shift-working surgeons-in-training from a geographically dispersed range of hospital sites across the region, and I felt that the logistical difficulties of arranging focus groups would likely jeopardise the rigour of the data, as it is highly likely that only a portion of the participants would have been able to attend the focus groups owing to their

fixed clinical commitments. It would, arguably, have been possible to conduct focus groups immediately at the conclusion of the CST course, when all the participants were physically in the same building. I felt that this would not have yielded appropriate data for my research questions, as there would have been no opportunity to return to the workplace following the training, so the phenomenon of skill transfer following training could not be explored and there would have been no time for the participants to reflect on and consolidate the experience of undertaking the CST intervention.

Sampling

All 14 study participants in the intervention group of the trial were invited by email to be interviewed. First and then second email reminders were sent to initial non-responders. Eleven participants were subsequently interviewed. Of the three participants who were not interviewed, one declined to be interviewed, one had moved abroad into a non-clinical job role and one participant did not respond to attempts at contact.

Setting

The majority of interviews (seven of eleven) took place in the research offices of University Hospital Coventry & Warwickshire. Three interviews were conducted by telephone, at the request of the participants, and one interview was conducted in the home town of the participant, as their preferred location. All reasonable efforts were made to accommodate the participants busy working schedules, and their preferences for location/interview modality. All interviews were conducted by me (HJ), and recorded using a digital voice recorder. All of the study participants were known to me as acquaintances and peers, and all interviews took place in a location familiar to the participants. As a fellow surgeon-in-training of similar age and professional experience, I had the advantage of matching the major social characteristics of the study participants, which previous research has shown is an important factor in determining the effectiveness of an interview (Sawyer 1995 as cited in Braun &

Clark⁴⁷⁷). It was easy to achieve a good rapport with all the participants at interview because of our similar backgrounds and the fact that we were already acquainted.

On the subject of reflexivity, I was approaching the interviews with my own considerable background experience of life as a surgeon-in-training. This meant that on the one hand I had the advantage of a well-developed understanding of the professional lifestyle, challenges and rewards of this career stage, but on the other hand I have my own formulated opinions on the current problems with surgical training, and of the value that simulation may bring to address them. I was present for the CST intervention, and the associated social events, and I designed, helped organize, and obtained the funding for the CST. I was therefore very personally invested in its success. In the conduct of this research, I was very careful not to impose my implicit biases on the participants, or of steering the interview based on them. I aimed to remain as objective and neutral as possible, and endeavoured to manage the duality of the participant-interviewer relationship sensitively and mindfully.

All participants were assured of confidentiality and gave their permission for the interview to be recorded and analysed for research purposes. A pre-piloted topic guide was used to bring some structure to the interviews as described above, and I used an open questioning style, except on rare occasions when I was seeking clarification on something. Participants were invited to freely add anything further they felt was relevant at the conclusion of the interviews.

8.4.3 Participants

The main demographic details of the interviewed participants are shown in table 31 below. All interviews were conducted six months after completing the CST intervention (March 2015). Written consent for participation was obtained at recruitment into the study, and had received the necessary prior ethical approval under the approvals in place for the trial.

Table 31. Demographic details of study participants

Study ID	Age	Gender	Months T&O*	Previous CST**
Participant 1	31	M	31	N
Participant 2	30	M	30	Y
Participant 3	29	M	22	N
Participant 6	26	F	2	N
Participant 7	29	M	4	N
Participant 8	27	M	6	Y
Participant 9	28	M	6	Y
Participant 10	29	F	34	N
Participant 12	26	F	5	Y
Participant 15	28	M	12	N
Participant 16	27	M	6	N

*Refers to the number of completed months of Trauma & Orthopaedic surgical training completed by the participant after graduation from medical school

**Any previous exposure to CST as a postgraduate marked as Yes

8.4.4 Analysis

I have used experiential thematic analysis to make sense of the data, and have structured my approach around Braun & Clark's checklist of criteria for 'good thematic analysis'⁴⁷⁸.

The digital audio recordings were transcribed by a professional typist, to a high level of orthographic detail, and I re-checked these against the tapes to be sure of accuracy. I transcribed two of the interviews myself as I believed it would be a valuable training exercise to gain this experience and understand the challenges of transcription. I only used the professionally transcribed versions for the analysis, to ensure consistency.

I undertook a process of initial reading and familiarisation with the transcripts, to gain an overall understanding of the data. Following this, I began the coding and searching for themes. I used a complete coding strategy to identify ‘anything and everything’⁴⁷⁹ of interest within the entire dataset. I generated a mixture of semantic and latent codes, with most being in the former category. This was a recursive process involving multiple revisions over a period of several weeks, until the entire dataset was completely coded.

The data coding process was comprehensive and each data item was given equal attention throughout the analysis. I used NVivo qualitative data analysis software, version 11.4.3⁴⁸⁰, to collate all relevant extracts for each theme. Once coding was felt to be complete, I searched for patterns within the coded data, from which to build themes.

Four candidate themes were initially developed from the coded data. Each theme had a distinct, central organizing concept⁴⁸¹ and two of these themes subsequently evolved to each have subthemes at a later stage in the analysis, to explain opposite positions within the domain of each given theme. A visual thematic map was developed during the analysis stage to enable exploration between the codes, themes and subthemes.

The themes that I generated during the analysis process were checked against each other and repeatedly referenced back to the original data set, to ensure they each had distinct scope and purpose, were faithful to the data, and that together they would provide a ‘coherent and meaningful’⁴⁸² overview of key concepts in the data that addressed my research question(s).

8.5 Findings (results)

Four distinct themes were derived from the dataset. These are outlined in table 32 below;

Table 32. Overview of themes and subthemes

Theme	Subtheme(s)
1: <i>'Getting into the operating theatre'</i>	Facilitative factors
	Obstructive factors
2: <i>'Factors driving learning from cadaveric simulation training'</i>	Intrinsic, surgeon-driven factors
	Extrinsic, environmental factors
3: <i>'Added value of cadaveric simulation training'</i>	
4: <i>'Professional transitions'</i>	

Each theme is discussed in turn.

Theme 1: Getting into the operating theatre

Traditional, on-the-job style training for surgeons involves observing and performing surgical procedures in the operating theatre. One of the main drivers for reform of surgical training, and provision of simulation training as a key part of these reforms, originates from concern that trainee surgeons are not being exposed to an adequate volume of operations during their training⁴. This is due to several factors that I have already discussed in this thesis including; reduced working hours^{32 86}, increased turn-over of patients and the burden of administrative, non-training tasks on surgeons-in-training^{70 89}. An in-depth discussion of these factors driving surgical training reform is presented in chapter 1.

On the theme of 'getting into the operating theatre', participants reported experiences that can broadly be clustered into 'facilitative factors', i.e. positive practices that enabled trainees to attend the operating theatre to gain training experience, and 'obstructive', negative factors that impeded access to training

opportunity in the operating theatre. These findings are summarized below in table 33.

Facilitative factors

Participants described training conditions they had experienced which were conducive to attending the operating theatre. These included advanced nurse practitioner (ANP) support with ward management of patients, which freed them up to go to the operating theatre;

“we’re supported by the ANP’s as well, so they fully understand...are very much on your side that, you know, when it’s your post-take day even though that’s the busiest day of the week pretty much, that’s the day that you get to the theatre and that should be the case” (Participant 7)

In addition, participants described having been allocated time-tabled, protected time within the working week to attend the operating theatre.

“We’re all allocated trauma lists pretty much at least once every other week, so there’s lots of operating” (Participant 9)

The experience of a facilitative working environment in which to attend the operating theatre was inconsistent between rotations and appeared to be dependent on location and senior support (or lack of).

“I think it really makes a difference with what senior team you have” (Participant 16).

References to ‘supportiveness’ were frequently made amongst participants that had experienced a facilitative environment for accessing the operating theatre. Support in this context was both practical in the form of, for example, ANP or other junior colleagues covering the ward work-load, and also psychosocial, whereby participants

were ‘expected’ to be present in the operating theatre and being involved in operations was an accepted part of their professional identity as surgeons-in-training.

The more senior participants reported that with increased experience their confidence grew in their ability to leave the ward to seek out training opportunities in the operating theatre. This was a combination of improved efficiency with performing ward-based tasks and a more pro-active approach to managing their own time, as well as a stronger sense of professional self-identity as a ‘surgeon-in-training’.

“I’ve been able to get down to theatre a lot more, just by, you know – leaving my F2’s and GP trainees [on the ward] and you know that actually, my training priorities are sort of different to them” (Participant 7)

Obstructive factors

Almost all the participants had experienced factors which had obstructed their ability to attend the operating theatre, and this dimension of the theme of ‘getting into the operating theatre’ was strikingly more prominent than that of facilitative factors in the dataset.

The administrative and medical workload on the ward, combined with a reported lack of junior staff, meant that participants were often missing training opportunities in the operating theatre.

Previously, the most junior member of the medical team would have been a postgraduate year 1 doctor (now called foundation year 1, or F1), whose principle responsibility would have been to manage the everyday medical and administrative tasks concerning the ward inpatients. These most-junior doctors have increasingly been moved away from T&O rotations into general surgical posts, as their working hours have been reduced, and as such the burden of managing the ward tasks in T&O in their absence have fallen to more senior surgeons-in-training, who would

have previously been free to attend the operating theatre and learning how to perform operations during this time.

The administrative workload involved in managing admissions, requesting tests, chasing results and processing discharges, which was traditionally the domain of the (now elsewhere) F1 doctors, is burdensome and time consuming for surgeons-in-training, whose working week is already shortened. This workload is such in part because of the increased patient turnover and pressure to progress discharges to manage capacity in the modern healthcare system, and partly because there is chronic underinvestment in technology in the NHS, and many of these processes are inefficient and unduly bureaucratic^{483 484}, involving technology that is outdated, slow and considered obsolete elsewhere⁴⁸⁵. It is striking that chronic process inefficiency is rather accepted as part of NHS working life in this population, and the lost productivity of the staff as a result is not addressed at a managerial level⁴⁸⁵. These administrative ward tasks carry no training value beyond F1 level, and serve to hinder surgeons-in-training efforts to meet their learning objectives for training⁷⁰.

In addition to the administrative workload, inpatients in T&O are often elderly, medically complex, and require a substantial input of ward care from doctors. Lack of staff at junior levels was cited as a key obstructive factor in accessing the operating theatre;

“The biggest problem [with accessing training in the operating theatre] is that we no longer have F1’s” (Participant 3)

“I actually found it very difficult to get training in theatre, leaving the wards was quite difficult” (Participant 10)

“previously what would have been a senior SHO [senior house officer, postgraduate year 3-4], able to be mobilized to theatre and clinic and get training experience is frequently the most junior person on the team...I spent 12 months... with no registrar and no F1 as a CT2...we’d at times have 50 [in]patients” (Participant 3)

“we’ve got people who are willing to train...it is just simply there aren’t the staff and junior cover required to cover the wards to facilitate those guys getting into theatre” (Participant 15)

There was a striking incompatibility between the curriculum (and expected educational achievements) of this group of surgeons-in-training, set against the demands of the everyday working environment they were in and the barriers that were preventing them from accessing appropriate learning opportunities in the operating theatre.

“ even when you are five years qualified...you are still essentially the most junior person on the ward delivering, catheterizing, cannulating, prescribing...that’s not what the CT2 [postgraduate year 4] curriculum is” (Participant 3)

From within this struggle to reconcile the challenges of being within a daily working environment which does not align with the training curriculum and what is required in terms of educational output, emerged a degree of competitiveness with peers for theatre access.

“Doing my core training I actually found it very difficult to get, uhm, training in theatre. I was competing with other SHOs” (Participant 10)

There was a tangible sense of frustration that accessing the learning opportunities to meet their learning objectives was such a struggle, and when set within the known landscape of low morale amongst junior doctors and attrition in trainee numbers within T&O, this is an important finding.

A consequence of the lack of appropriate training opportunities for this group meant that when they progressed to the next level of training (known by various synonyms including; ‘first year registrar’, specialist training year 3, ‘ST3’), these surgeons-in-training had not reached the operative independence of their predecessors. They were therefore unable to offer any peer-training to their more junior colleagues, and were needing to undertake the ‘index’ operations themselves that traditionally would’ve been the domain of the more junior surgeons-in-training, as they had not

yet had the exposure at an earlier, more appropriate career stage, because of the reasons outlined above. Therefore, it was apparent that the problem perpetuates downstream;

“Just the number of people around us is one thing, especially as a lot of the reg’s [registrars] being quite junior” (Participant 6, on discussing barriers to accessing training)

“The junior reg’s [registrars] still need those index procedures” (Participant 6)

“I had an ST3 [as registrar] at the time so I didn’t get to do much” (Participant 7)

Hence surgeons-in-training who work in a unit without adequate junior ward staff/ANP support, and whom are in a team with an ST3 registrar, face considerable challenges in achieving the exposure to appropriate training opportunities to meet their required educational outcomes.

A further barrier to accessing training was the balance of cases encountered in the training post – participants who were working in the large major trauma centres reported that a predominance of complex, poly-trauma patients within their typical caseload meant that there was a lack of opportunity to practice the basic procedures which they were required to achieve competency in. These complex procedures, whilst interesting, were consultant-led and did not provide a training opportunity that was appropriate for their level.

“being a level one trauma centre...taking up [time on the trauma list] with a big case like a spinal case or a pelvic case and so my experiences of normal, routine, trauma operations is actually fairly limited” (Participant 10).

Once training was taking place in the theatre, there were other obstacles to learning, including a perception of pressure; both in time, and professional pressure from other members of the theatre team who might be impatient that a trainee surgeon was performing the operation.

“There’s a lot of pressure, I think, on the trainee and the consultant to try and hurry up the case and get it done quicker” (Participant 15)

“..external pressures, you know, the anaesthetist or something like that or an unwell patient” (Participant 7)

In addition to the barriers discussed in relation to the daily working environment, the regulatory change of the doctors’ working hours through the introduction of the EWTD has negatively influenced the participants’ experience of accessing training. A shorter working week inevitably reduces the opportunities for training, and some participants reported seeking out training opportunities outside of their working hours.

“I think there’s a lot more pressure now to spend your own free time to get the training that you need...that is sort of what you have to do... it is much more of a problem than it used to be” (Participant 15)

Table 33: Factors influencing surgeons-in-training access to learning opportunities in the operating theatre

Facilitative Factors	Obstructive Factors
<p>Colleague support;</p> <ul style="list-style-type: none"> • ANP's • Junior doctor peers • Senior support <p>Protected time;</p> <ul style="list-style-type: none"> • Timetabled • Trainee allocated lists • Designated free from other duties <p>Trainee seniority;</p> <ul style="list-style-type: none"> • Improved efficiency with non-training tasks • Prioritize own learning • More proactive in seeking training • Assertiveness to leave ward/delegate 	<p>Lack of junior staff covering wards</p> <ul style="list-style-type: none"> • Loss of F1 doctors • Rota gaps • Shorter working week <p>High ward workload;</p> <ul style="list-style-type: none"> • Medically complex patients • High patient turnover • Inefficient bureaucratic processes • Not educationally valuable for this group – could (should?) be done by others e.g. ANP's • Mismatch between curriculum and actual working environment <p>Wrong case-mix;</p> <ul style="list-style-type: none"> • Too specialized • Not enough routine trauma exposure <p>Time pressure in theatre;</p> <ul style="list-style-type: none"> • Unwell patients • Pressure from non-surgical staff

Theme 2: Factors driving learning from CST

The theme of factors driving learning from CST provides a detailed insight into how, and why, surgeons-in-training learn skills in a CST intervention. Participants'

experience and opinions on these can be broadly separated into two sub-themes; intrinsic, surgeon driven factors, and extrinsic, environmental/course-related factors.

Intrinsic, surgeon driven factors

Self-perception of operative confidence influenced learning following CST. Those with low-self reported confidence in their operative skill appeared to make the greatest gains following the CST intervention;

“For me before I started the course, hemiarthroplasty [one of the taught index procedures] was my Everest and I think after that course I was much more confident in approaching it” (Participant 12)

The ability to ‘buy into’ the simulation exercise and behave as if operating on a real patient was important for optimal learning and the ability to do this successfully varied between individuals. One participant struggled with ‘suspension of disbelief’ in the low fidelity setting and appreciated the value of the high-fidelity of the CST in achieving an immersive experience;

“so in a [low-fidelity simulation setting] workshop, it is much easier to slip into not quite doing it properly, such as a soft tissue guide, because there isn’t soft tissue to worry about...suddenly the whole illusion breaks down” (participant 8)

An ability to push one’s own learning boundaries within the safety of a simulation exercise enabled learners to gain maximal benefit from the CST intervention. For example, whereas in real-life an inexperienced trainee often seeks reassurance from the trainer before progressing through each stage of the operation, in the simulation exercise the trainee can move beyond their comfort zone more confidently and improve their operative fluency without the risk of causing patient harm.

“I got to the end and I thought ‘Ah I haven’t actually asked for help’ as supposed to in a real situation...I will be constantly looking for reassurance that the guide wire

is in the right place, that you know, everything was, everything was right...and almost asking permission before going onto the next step” (Participant 6).

Correctly timing the delivery of the CST intervention within training was perceived to be a crucial intrinsic, learner-dependent factor to its success. Most of the participants felt that delivery of the course in its original format was best suited to the beginning or middle of the second year of core training (CT2, equivalent to postgraduate year 4). This was because at CT2 level, the participants did not have much independent operative experience, had experienced many (if not all) of the obstacles to accessing the operating theatre discussed above, and competence in these index procedures was expected to be achieved to progress on to registrar training. This was a consistent finding across participants at CT2 level who felt the course was appropriately timed for their stage of training, and amongst participants at a higher level (ST3) who felt the course would have been of greater benefit to them a year earlier.

“I felt that the level I went into it at the beginning of CT2 was perfect because you, you’ve had a bit of time in trauma theatre, but maybe not as much independence as the senior trainees...and then getting a lot of confidence from having seen the four index procedures [on the CST course] (Participant 16)

“Middle of CT2 [would be ideal], those procedures are, apart from the fasciotomies, I think they are all essentials for becoming a registrar” (Participant 2)

Timing of delivery with respect to the commencement of a trauma rotation was also raised, with the course being more useful if it were to be made available before the placement begins (as compared to 6 weeks in, as it was in this study), and that there was a perceived risk of learned skill attrition if the course was delivered too early.

“If I had it just before [trauma] that would be even better...you need to be doing it relatively fresh, within four to six weeks” (Participant 3)

Extrinsic, environmental factors

Participants were paired during the CST training intervention, and whilst one was operating as ‘first’ surgeon, their partner was assisting or acting as scrub nurse. The paired-learning nature of the CST intervention was perceived as valuable, as there was the opportunity to learn from the experience of a colleague partner;

“I think having...two participants working together was very useful because you see one [procedure], your colleague doing it and then you do it yourself, you can kind of learn from each other and even if you do make any pitfalls you can kind of learn from that experience, and think what you would have done differently and so I think learning from each other is a really good thing as well” (Participant 15)

The multi-disciplinary environment of the CST intervention was perceived as valuable and enhanced learning, both in terms of enabling dialogue about their performance between allied health professionals (scrub staff and radiographers), and through the opportunity to assume the role of scrub-nurse. This gave participants an insight into the role the scrub-nurses play in ensuring the smooth progress of an operation, and furthermore to improve their own knowledge of the sequence of steps in a procedure, as the scrub-nurse is required to anticipate the next stage of a procedure and have the correct instruments to hand;

“Actually what I found very helpful was being a scrub nurse and watching and anticipating and giving them the next thing and the next thing” (Participant 8)

“I liked the way that we had the theatre staff come in as well, it was really good to get their opinions on things...I think that was very good” (Participant 10)

“...your interaction with your team, having that one on one feedback as well as you were doing the process [was beneficial to learning]” (Participant 16)

One of the key extrinsic features of CST that drives learning is the opportunity to perform operations in their entirety as first surgeon, and in an intensive way; i.e.

several successive operations in a short overall period of time. This would not typically be encountered in the real-life operating theatre environment, except for the most senior trainees. The participants valued this opportunity and felt that it enabled them to progress their skills more quickly than usual;

“[CST] gives you an opportunity to do a lot of operations in a short period of time as first surgeon” (Participant 6)

“I think [the CST course] was the first time I had ever done a hemiarthroplasty entirely on my own” (Participant 12)

Another key feature driving learning was the intensive nature of the supervision during the CST intervention. Each operation was supervised by a consultant, who provided real-time feedback and guidance, and were given the scope to challenge trainees (within the domain of the curriculum) as they judged appropriate. The consultant faculty were self-selected, they had volunteered to teach on the course, and were thus naturally enthusiastic and invested in teaching trainees.

“To have high quality teachers one-on-one was fantastic... [the CST course was] an excellent way of learning, having the consultant stood over your shoulder which is something you might not have in the actual theatre itself” (Participant 3)

“[Amongst] the things I found most helpful [about the CST] was you were getting one-on-one consultant level teaching”

Having intensive supervision also helped the learners maintain the fidelity of the simulation, as the faculty helped maintain the illusion that they were in the ‘real’ operating theatre through their non-verbal cues and behaviours;

“It keeps you switched on and stops you lapsing, so [you aren’t] doing the ‘oh in real life’ I would have done this’, that’s really helpful” (Participant 8)

The consultant faculty were allocated stations according to their sub-specialist interests, and their super-expertise was valued by the participants;

“having an ankle surgeon [specialist] at the ankle station was good, because ankles can always be a bit fiddly and people have certain tips and tricks, it was very helpful” (Participant 15)

Value beyond technical skills training

Cleland et al⁴⁷⁴ ran a rapid ethnographic study of two surgical boot-camp training courses which were delivered to early stage-postgraduate surgical trainees. The study aimed to understand the socio-cultural influences of this intensive training and the wider implications for simulation based-education. The authors found that intensive boot-camp style training (of which CST is a variant) is ‘as much about social and cultural processes’ as it is about ‘individual, cognitive and acquisitive learning’⁴⁷⁴. These findings build on previous work examining how surgeons-in-training ‘become through doing’. Prentice, in an ethnography examining how medical students and junior doctors learn surgical skills in the operating theatre, states that in order to gain a full understanding of how a ‘resident comes to embody the knowledge, skills and values of a surgeon requires understanding how social milieu and guided practice interact’⁴⁸⁶. Prentice describes the ‘guided physical training in the operating room’ as embodying the ‘technical and social lessons of surgery’, even where the skills being taught are purely ‘technical’. Her findings reported that technical skill is only “20 percent” of the overall skillset required of a surgeon, with unspoken ‘tacit’ knowledge, clinical judgement and moral behaviour forming a substantial part of what is required of a surgeon, beyond technical proficiency⁴⁸⁶.

Previous qualitative work examining surgical practice has ‘shown surgical action in detail, but have little to say about how surgical trainees learn to fit themselves into the team, how they take on increasing levels of responsibility and how they develop the moral qualities of a surgeon’⁴⁸⁶. Prentice attempts to ‘unpack the unspoken lessons of surgery’ by framing it within Bourdieu’s (Bourdieu 1977 cited by Prentice⁴⁸⁶) discussion of the ‘symbolically structured environment’. According to Bourdieu, the ‘structures of an environment build particular organising principles, habits and the ways of being into the minds and bodies of cultural actors’ (surgeons

in this instance), and the symbolic structured environment (the operating theatre) exerts an anonymous, pervasive, pedagogic action'⁴⁸⁶. Prentice elaborates on this with reference to surgical training, that in the operating theatre the highly ritualized 'space, time and costuming provide structuring effects that make the imitation of a surgeon's actions and attitudes have meaning'⁴⁸⁶. This helps create a positive economy in learners by instilling 'the social hierarchy of the operating room', a hierarchy that 'places surgeons in the centre, and gradually (with increasing experience) moves surgical residents into full participation at the centre of the action'.

The value of CST, therefore, goes beyond provision of technical skills training. The ultra-high environmental fidelity of CST replicates the 'symbolic structured environment' of the operating theatre and allows the learner to begin to become socialized in the practice of surgery and to 'become through doing', learning both technical and non-technical skills in a highly realistic environment, which itself exerts a 'pervasive pedagogic action'.

The course was organized very carefully to maximize training opportunity, and this was recognized and appreciated by some of the participants, who felt the organisation was an additional extrinsic factor that influenced their ability to learn from the CST

"I think it [the CST intervention] was really well thought out" (Participant 10)

There was recognition that simulating the complexities of the real-life operating theatre was extremely difficult, and that CST was the best available simulation modality to try and achieve this replication;

"It is very hard to simulate training in orthopaedics...certainly cadaveric training is probably the only way you're going to be able to do that" (Participant 2)

Some aspects of the CST intervention compromised the fidelity of the training experience, in particular the specimens moved around during the hemiarthroplasty procedure (normally the patients' body-weight and positioning aids prevent this in

real-life). We had waist-to-toe-tip specimens rather than whole cadavers, and so these obviously weighed less than a whole body.

“for the hemiarthroplasty they [the cadavers] were just moving around a little bit and it wasn’t as realistic as you had in theatres” (Participant 16)

This also negatively affected the realism of re-locating the hip once the implant was in position for similar reasons;

“Having just the one leg for hemiarthroplasty made setting up quite difficult and when you tried to relocate the hip” (Participant 10)

One of the participants felt that the bone quality in the specimens was low which compromised the fidelity of the simulation and generated additional surgical complications, which perhaps were not appropriate for the junior level of the participants. Poor bone quality ‘osteopenia’ is often encountered in the real-life operating theatre, and adds a challenging dimension to the successful performance of an operation.

“The bone wasn’t of fantastic quality...and complications, shall we say, arose”
(Participant 12)

The factors driving learning from CST are summarized below in table 34.

Table 34. Factors driving learning in CST

Intrinsic, surgeon-driven factors	Extrinsic, environmental factors
<p>Self-perception of skill level;</p> <ul style="list-style-type: none"> • Least confident participants appeared to make the most gains <p>Willingness to ‘buy-in’ to the simulation;</p> <ul style="list-style-type: none"> • Suspension of disbelief • Staying in ‘character’ <p>Pushing boundaries;</p> <ul style="list-style-type: none"> • Move out of comfort zone • Not ask for help as often • Safe space to make mistakes <p>Timing of delivery;</p> <ul style="list-style-type: none"> • Participant at correct stage of training for maximal benefit – early/mid-CT2 • Do CST just before start of trauma rotation 	<p>Paired learning;</p> <ul style="list-style-type: none"> • Practice assisting • Learn from colleague <p>Multi-disciplinary simulation;</p> <ul style="list-style-type: none"> • Experience being scrub nurse • Alternative perspectives <p>Perform operations in their entirety;</p> <ul style="list-style-type: none"> • Fluency • Momentum <p>Intensive Consultant supervision;</p> <ul style="list-style-type: none"> • One-to-one • Real time feedback • Super-specialised • Tips/tricks • Unusual approaches <p>Becoming through doing;</p> <ul style="list-style-type: none"> • Non-technical skills • Highly realistic space, time and costuming • Sociocultural, ‘unspoken’ lessons of surgery • Stirring to practice

Theme 3: Added value of CST

The theme of ‘added value of CST’ was the unifying concept of a cluster of findings relating to how cadaveric simulation was valuable in a manner beyond that of other simulation modalities. Clearly, with cadaveric simulation being so much more expensive to deliver than other lower fidelity, inorganic types of simulation, it needs to bring additional benefits that these cheaper alternatives do not, to justify the additional cost. Participants were asked if there were features of the CST that were particularly useful to them in developing their skills as surgeons-in-training, and whether they felt CST offered value beyond that of other simulation modalities.

Anatomical fidelity

Anatomical fidelity, the presence of a soft-tissue envelope and neurovascular structures as seen in life, were reported by participants as valuable features peculiar to CST that could not be found elsewhere. Intuitively, deceased human bodies offer the most realistic representation of living anatomy, which is extremely difficult, if not impossible, to replicate by other means. Where visual representation of anatomy can be achieved by use of sophisticated computer and virtual reality programming, the haptic, tactile feedback and ‘tissue tension’ experienced in cadaveric simulation is unparalleled.

“Tissue tension is something that is quite unique to the cadavers really” (Participant 8)

“Technical skills, in going through [dissecting] various layers, you can’t simulate [that] with dry bone” (Participant 7)

“Life-like I suppose, with tactile feedback” (Participant 1)

“I think in terms of how high fidelity it was compared with what you normally do [in real life], it was very close” (Participant 16)

The presence of a soft tissue envelope made the educational experience much more valuable for participants as they had to navigate the neurovascular hazards as they would in real life, and they couldn't 'cheat' by obtaining direct visualisation of the bone, as is possible in low-fidelity benchtop models such as sawbones. They were therefore more invested in the authenticity of the simulation experience, which became immersive, and led to other cognitive benefits.

Within the immersive-ness of the experience, participants 'bought into' the realism and began to behave as they would in the operating theatre. This investment in the simulation exercise revealed another important area of the added value of CST – consequence and patient safety. Participants, whilst immersed in the realism of the training experience, knew that there were no real consequences to their making a mistake, and there was no risk to patient safety. In treating the exercise as a 'full dress rehearsal' for real life operating, the participants felt confident to push their own boundaries, and progressed their learning as a result.

"It doesn't matter if you get it wrong" (Participant 3)

"I think it's good for the trainee, because they go through all the steps [in CST], they make sure they feel happy and confident, they've gone through the motions, and they can consolidate that on a cadaver first...and it is good for patients because they get someone [a surgeon], they're not practicing on a real person" (Participant 15).

Skills learnt in the CST intervention were directly transferred to the operating theatre, as result of this 'dress rehearsal' opportunity. There was no need for the participants to aggregate or embellish their learning before taking it to the operating theatre, the learning from CST was whole, or complete. This finding is in contrast to the learning gains from low-fidelity simulation, where learning is necessarily restricted to particular component skills under instruction, and cannot be automatically, or easily, be scaled up to the real life operating theatre.

For example, one participant described how a week after attending the CST course, they had been on-call over the weekend and been asked to perform a lower-limb

fasciotomy. This is an emergency procedure, performed when intra-muscular pressure is dangerously raised (for example as a consequence of a fracture), and the muscle compartments need to be opened surgically to release the pressure. The participant describes how there was no-one available to supervise them performing the procedure, but having completed the procedure on the cadaver during the CST course the previous week, they felt confident of the surgical landmarks and hazards, in a way that reading about the techniques in a textbook would not have achieved.

“I knew where the perforators would be, it all went very smoothly, and it is one of those things where if you’ve never done something before you can read it in a book, but you’re not going to be sure of yourself, and I think that once I’ve done something [on the cadavers] I know I can do it, then I’m just much more happy and confident that I’ll be competent to do it...doing that on a specimen rather than a person, in that situation [the fasciotomy] especially because it is not something that you see every day, I might not do one for another few years, it is very valuable” (Participant 1).

Another participant described how their supervising consultant knew that they had recently successfully completed a fasciotomy procedure during the CST intervention, and so when a real case was encountered a few weeks later, the consultant was happy to let the surgeon-in-training perform the entire operation on the patient, confident in the knowledge that the trainee had previously achieved competence in the simulated environment.

“I definitely used what I learnt [on the course] ...it’s definitely made a difference to training” (Participant 16)

Safely pushing boundaries

Within this sense of safety, and yet investment in the environmental fidelity of the simulation, participants reported valuing the opportunity to have the time to perform the operation in its entirety, without the usual time pressures of the operating theatre;

“so that [the CST intervention] was a great place to just do a procedure and not be cut there, cut there, cut there, just crack on, do it and it doesn’t matter if you get it wrong” (Participant 3)

“in an environment that wasn’t time pressured or, no external pressures, the anaesthetist or something like that or an unwell patient...was very helpful indeed and from a learning point of view I think I learnt more in the cases I did there [the CST intervention] than the vast majority of cases that I do in theatre” (Participant 7)

Participants valued the real-time feedback that consultant trainers were able to provide during the CST, and the opportunity to complete workplace based-assessments from their operative performances;

“I really liked the one on one consultant feedback. I thought it was a really nice way of doing things” (Participant 16)

“It is quite hard in day to day training sometimes to get the consultants to sit down and do the forms properly and give you constructive feedback. Often they will say ‘oh just fill it in and send it to me’, but they had to do it properly [during the CST intervention]” (Participant 6)

In addition, participants valued the opportunity to be taught ‘tips and tricks’ from the consultants during the CST, and to have the opportunity to practice some of the more unusual surgical approaches, for example the posterolateral approach to the hip.

“People have certain tips and tricks...it was very helpful” (Participant 15)

“For me, the best bit about cadaveric training is getting to do things [approaches] that you don’t normally do” (Participant 2)

CST is of most value after learning basic concepts using low fidelity simulation

The timing of CST with respect to delivery of other low-fidelity simulation training opportunities was also explored with participants, with a particular emphasis towards understanding whether CST has an adjunctive or replacement role when compared with low-fidelity simulation.

As a pre-requisite to completion of CT2, all trainee-surgeons must undertake the AO (Arbeitsgemeinschaft für Osteosynthesfragen) Foundation Basic Principles of Fracture Management Course⁴¹². This is an interactive course which teaches the basic concepts of stability, physiology of bone healing, and reduction and fixation techniques for simple fractures using low-fidelity, plastic bone simulation. Given the consensus amongst participants that the CST intervention was best delivered in the middle of the CT2 year, i.e. within the accepted timeframe for AO course completion, it was interesting to explore whether they felt that CST carried most educational benefit when delivered before or after the AO course.

Participants reported that the CST was most beneficial after they had grasped the basic principles of fracture management via the AO course, and that the more sophisticated simulation environment of CST allowed them to build on what they had already learnt in the low-fidelity environment. The AO course is very valuable as a first introduction to the principles and surgical instrumentation, which do not require the expense of CST to impart to surgeons-in-training.

“dry bones are really good, I think for basic principles and just getting familiar with technique and equipment...and then being able to apply those basic principles [in CST]..If you like, a higher level of simulation” (Participant 7)

Participants were also asked about their opinions on whether there is a role for embedding CST within the curriculum, making it routinely accessible to surgeons-

in-training as part of their formal teaching programme. The response was very much in favour of this approach;

“Absolutely. Absolutely, I think it’s the way forward really” (Participant 15)

“It’d be a big loss if you weren’t able to build it into the curriculum” (Participant 16)

There was strong feeling that CST should be centrally funded and provided free to surgeons-in-training;

“They [the courses] should be delivered in region, for free to your trainees at the appropriate time” (Participant 3)

There was a sense that CST was a tremendous and yet presently underutilized training resource, and that the opportunity to receive CST was regarded as a privilege;

“I think having cadaveric [simulation] training is an unbelievable privilege for us and really, really useful” (Participant 10)

“I thought it [the CST intervention] was perfect, it was fantastic and I’m so lucky to be part of it” (Participant 12)

Table 35 summarises the features of CST which cannot be obtained through low-fidelity simulation.

Table 35. The added value of CST

Superiority of CST as compared other simulation modalities
Anatomical fidelity; <ul style="list-style-type: none">• Soft-tissue envelope• Neurovascular hazards• Tissue tension• Haptic feedback
‘Dress rehearsal’ for real surgery <ul style="list-style-type: none">• Whole learning, enabling direct skill transfer to the operating theatre• Can build on and refine foundational skills learnt in low-fidelity simulation• Becoming through doing, ‘stirring into practice’
Offers solutions to some of the barriers to learning in the operating theatre; <ul style="list-style-type: none">• No time pressure• No pressure from anaesthetist/other theatre staff• No risk to patients

Theme 4: Professional transitions

The theme of ‘professional transitions’ encompasses the findings related to participants’ experience of undergoing the significant change in role and responsibility when they begin registrar training, and the role of CST in offsetting the challenges associated with this transition.

SHO and Specialist Registrar roles and responsibilities are very different

Medical training grades have undergone several iterations of renaming and reorganisation (as discussed in chapter 1), yet the established hierarchical structure of doctors-in-training in the workplace remains unchanged, and, colloquially faithful to the ‘old system’. The first-postgraduate year doctors (called ‘F1’ or ‘foundation year 1’) are still known as ‘house officers’, post-graduate year 2 to 4 doctors (called ‘F2’ and core-trainee 1 and 2 ‘CT1/2’) are known as ‘senior house officers’ (most commonly abbreviated to ‘SHO’) and postgraduate year 5 and beyond are known as ‘specialist registrars’, commonly abbreviated to ‘SpR’.

The distinction between these professional identities (house officer, SHO and SpR) have both a practical and more abstract, philosophical dimension. These practical differences are relatively straightforward to define by reference to the curriculum and educational objectives of each role. The overall remit and level of clinical responsibility is different between each, and each role is distinct within the ubiquitous hospital ‘on-call’ system. What is more difficult to elucidate is the more abstract difference; how these distinct levels are viewed within the culture of a department or specialty and the unspoken expectations placed on these roles, along with individuals own sense of professional self-identity within these roles, and how this changes with upward progression.

Arguably, the practical difference within SHO grades (F2/CT1/CT2) is not particularly large (in T&O at least); all spend most of their time caring for patients on the ward, admitting patients and handing administrative tasks. The more senior

SHO's may have greater access to the operating theatre (notwithstanding the issues with access discussed above), but they can quite reasonably be considered a fairly homogenous group both in terms of job remit and professional identity.

A significant transition for surgeons-in-training occurs between postgraduate year 4 (CT2) and year 5 (ST3), when they go from being an 'SHO' to a specialist registrar, 'SpR'. This is a very important, and a much larger, transition than across the HO/SHO years for several reasons. From a practical standpoint, the day-to-day job of the SpR is very different. Where the SHO is concerned mainly with managing ward inpatients, from an administrative and medical standpoint, the SpR is in the outpatient clinic (sometimes in sole charge), seeing a high turnover of patients and making decisions on their management from a surgical point of view. They can (within reason), delegate administrative tasks to more junior colleagues. The SpR will have dedicated, timetabled operating lists, including emergency trauma admissions and elective, planned procedures, (varying according to the rotation). They may, appropriately, and sometimes inappropriately, be left alone in the operating theatre to do the list, and will invariably be the most senior surgeon on the hospital premises out of hours. In addition to the practical differences in job role, there are some unspoken, less tangible differences that signify the progression to SpR level. The SpR will dress differently, not carry the ubiquitous 'junior doctor' paraphernalia necessary for ward management of patients such as a stethoscope, and will be professionally addressed as 'Mr/Mrs/Miss' rather than 'Dr' – a status indicator of having passed the membership exams of the Royal College of Surgeons.

There is something of a paradox therefore, that in a climate of time-based, linear (and in some cases, 'run-through') continuum training, that such an arguably archaic distinction between the two roles is still in evidence. This transition between roles also happens, rather arbitrarily, during the first Wednesday of August ('Black Wednesday'⁸⁰). A surgeon-in-training will find themselves comfortable and competent in their role of senior-SHO on the Tuesday before the changeover, and then quite literally overnight will assume the very different role of SpR – and bear the weight of the expectation of colleagues and the 'system' that they suddenly be fully fledged and competent for the considerable demands of the SpR role. They may at once find themselves in the role of first surgeon in the operating theatre alone and

unsupervised for the first time in their career, and are likely to be the most senior grade of surgeon on-site out of hours.

Study participants described their experiences of the transition from SHO to SpR. Due to the heterogeneous nature of the training level of the participants (table 31), some were discussing this change in anticipation of undergoing it in the near-future, and some were reflecting on their experience of having recently made the transition to SpR.

“I found it a massive step when I did it over there [a district general hospital] so I worked as an SHO for three months in their T&O department and then worked as a junior registrar and did find it an enormous step” (Participant 2)

Large tertiary healthcare centres have been described as having more ‘defined physical and socio-cultural silos’ when compared to smaller hospitals, where the inter-professional working relationships are more fluid and less hierarchical⁴⁷⁴, which can manifest as less structured provision of senior support in the clinical workplace.

The level of consultant supervision that could be expected at the start of the first SpR rotation had an impact in helping offset the transition;

“Because I know how well supported the ST3’s are at [regional teaching hospital] I really don’t have any anxiety, we won’t be left to do something inappropriate”
(Participant 8)

“I think they give you more responsibility with less supervision over there [district general hospital] than they do over here [major teaching hospital]” (Participant 2)

CST can help manage the SHO to SpR transition

Clearly, in a situation where surgeons-in-training are not getting appropriate training in surgical skills due to barriers in accessing the operating theatre during their SHO years, there is a considerable patient safety implication of ‘unpreparedness’ for the SpR role. This transition is therefore of great interest from an educational research perspective because it is ripe for improvement to facilitate the process and mitigate any potential risks to patient safety.

CST, therefore, may have an important role to play in enabling this group of surgeons-in-training achieve competence in index surgical procedures before making the transition to SpR, and provide an adjunct to training in the operating theatre, which may be of particular value in a climate of restricted access.

Jensen et al⁴⁸⁷, in a multi-site ethnographic study into learning in the operating theatre, describe surgical training as a process of being ‘stirred into practice’, which they define as learning by ‘participating in the practice of providing high quality-care’, where the aim is ‘to teach students to be surgeons instead of teaching them to perform surgery’. They describe real-world surgical practice as consisting of three dimensions; cultural-discursive arrangements: ‘sayings’, material-economic arrangements: ‘doings’ and social-political arrangements: ‘relatings’⁴⁸⁷.

CST as ultra-high fidelity training is unique within the field of simulation-based education in that it can contain all three of Jensen’s constituent dimensions of surgical practice in one training package, and participation in CST could be seen as a means of obtaining ‘social capital’, gaining not only skills and knowledge (‘sayings’ and ‘doings’, in this analogy), but also ‘insider information’ (‘relatings’) on how to behave like a surgeon in the operating theatre environment⁴⁷⁴.

Therefore, CST potentially has an important role to play in enhancing the preparedness of surgeons-in-training, not just for the markedly increased operative responsibility and skill expectation that comes with the SpR role, but also to enhance the process of ‘becoming a surgeon’ and being ‘stirred into practice’⁴⁸⁷.

The ‘constrained’ nature of intensive surgical training courses (such as the CST intervention) has been described as an ideal ‘micro-environment’ in which to study the ‘socio-cultural processes involved in becoming a surgeon’⁴⁷⁴. I did not explicitly consider the socialisation aspects of the CST intervention in the interviews, but I was present for the entirety of the intervention, course-refreshment breaks and ad-hoc, trainer-trainee social interactions as a non-participatory observer, and my observations and field notes have enabled me to gain a greater understanding of this phenomenon beyond that which is immediately apparent from the interview transcripts. Intensive simulation-based surgical training courses such as the CST intervention are ‘an inherently social activity as they bring together groups of trainees and faculty away from the everyday clinical environment’⁴⁷⁴, and can serve to facilitate the socialisation of trainees into the etiquette of surgery.

The participants reported that CST had a positive role in managing the transition, and that they had directly transferred skills learnt in the simulated environment into the workplace;

“I actually had to do that [lower-limb fasciotomy] for the first time a few days after the course so it came in really handy. I was sort of by myself, having to do that [the operation, as a new registrar]. So it [the CST] was a bit of a lifesaver actually”
(Participant 1)

Improvement in operative confidence, as a direct result of undertaking CST, was a recurring concept in these experiences;

“Getting a lot of confidence from having seen [and performed] the four index procedures and used that to then to build on in trauma theatre” (Participant 16)

[On making the transition to ST3] *“I think it will significantly increase the confidence of having done [the index operations at CST] essentially”* (Participant 7)

8.6 Discussion

8.6.1 Context of findings within existing research

Cleland's ethnographic study⁴⁷⁴ of a UK based simulation 'boot-camp' for early career stage (CT1) surgeons-in-training, described three broad areas of educational gains following the training; technical (and non-technical) skills, 'cultural capital', which the authors describe as 'resources in the form of learning what knowledge, skills, and values were needed to succeed in the surgical training system' and 'social capital', in terms of extending their mentoring network⁴⁷⁴. The authors acknowledge finding evidence of a distinction between the explicit and 'hidden' curriculum within the boot-camp environment, with the latter adding value to the training by facilitating 'enculturation and socialisation into surgical training'. They state that because this intensive training environment supports 'both formal skills learning and informal learning about how to be a surgeon through social and cultural processes', it is important that simulation-based training programme developers and researchers 'address the social and cultural aspects of learning when planning similar enterprises', as educational interventions do not occur in 'social, historical or cultural isolation'⁴⁷⁴.

Their conclusion aligns with the findings of Jensen's⁴⁸⁷ study of how medical students learn in the operating theatre, that the phenomena of surgical learning can be perceived as 'instances of transformation in and among social practices', that students learn by 'participating in the practice of providing high quality care', beyond simply technical skill acquisition in isolation, and the overall aim therefore is to teach 'students to be surgeons instead of teaching them to perform surgery'⁴⁸⁷.

My analysis shows that the socio-cultural features of CST were valued by participants, it helped with preparing them for the SHO-SpR transition and developing their professional identities and confidence as surgeons, and that the ultra-high fidelity nature of the simulation had additional, nuanced 'cultural capital' benefits beyond the more obvious remit of developing technical skill acquisition. These benefits are unique to CST, as a consequence of successfully replicating the

symbolically structured environment of the operating theatre, with its associated ‘pervasive pedagogic action’ in developing both technical and non-technical ‘tacit’ social skills and knowledge⁴⁸⁶.

8.6.2 Strengths and limitations

This study has several strengths. In undertaking this piece of research I have followed Yardley’s ‘open-ended, flexible’ quality principles (Yardley 2000 cited by Braun & Clarke⁴⁸⁸), which ‘represent one of the most successful attempts to develop theoretically neutral validity criteria in qualitative research’⁴⁸⁸. I have demonstrated ‘sensitivity to context’ by contextualizing my research within the literature and explicitly declaring my ontological and epistemological assumptions, and by consideration of the participants’ perspectives, my influence in the role of researcher and peer-colleague and the associated ethical issues.

Commitment and rigor has been demonstrated by a thorough data collection and analysis procedure, with appropriate methodological competence as researcher, and in-depth engagement with the research topic on both a professional and personal level. Transparency and coherence are central to robust qualitative research practice, and I have endeavoured to show this by presenting a clear analysis that is faithful to the dataset and theoretical framework, and reflexive in acknowledging the role that I have had, both as researcher and peer-colleague of the participants, in shaping the research. This is the first qualitative study (to my best knowledge) on the role of CST for training surgeons and thus adds to the existing evidence base in this area.

This study also has several weaknesses. The participants were all from one training region in the UK (West Midlands), and were individuals who had agreed to take part in the educational trial, and thus might represent a particularly motivated cohort of surgeons-in-training who are interested and engaged in simulation training research.

The interviews were conducted six months after the CST intervention, and represent the participants’ experiences in training at that point in time. Ideally, if resources had permitted, it would have been helpful to have repeated the interviews at a later stage,

to further explore the longitudinal nature of the impact of CST, and to also interview the consultant trainers involved in delivering the CST, to gain an understanding of their perspectives.

I also recognize that three of the interviews were conducted over the telephone, which may have shaped the data obtained from these participants, as ‘virtual’ interview methods such as this can mean that information can be lost or misconstrued in the absence of non-verbal cues that occur during the interaction of a face-to-face interview⁴⁷⁷. The telephone interviews were shorter in length than the face-to-face interviews, and the data generated is likely to be different than that from a face-to-face encounter. That said, a telephone interview with these three participants is arguably better than no interview at all, as these participants were unable to commit to a face-to-face appointment due to their workload and location.

8.7 Conclusions

Within the study population, CST is highly valued by surgeons-in-training, and improves confidence and preparedness for assuming the role of Specialist Registrar. CST can help offset the issues around accessing conventional training opportunities in the operating theatre in the early-stages of training and it may serve to help surgeons-in-training achieve competency in index surgical procedures more quickly, with resultant patient safety benefits.

CST is advantageous over low-fidelity simulation, and offers ‘added value’ for several reasons. The ultra-high anatomical fidelity of cadavers presents an opportunity to practice operations with an unparalleled realism. The intensive consultant supervision with real-time performance feedback, and opportunity to perform an operation in its entirety as first surgeon allows learners to push their own boundaries within the safety of a simulated environment. The multi-disciplinary nature of CST allows the learner to experience the perspective of scrub nurse, which

enhances their knowledge of sequencing of operative steps, teaches anticipatory skills and team-working.

CST has an adjunctive role alongside conventional surgical training and low-fidelity simulation. The middle of the CT2 (postgraduate 4) year was reported to be the best time to deliver this training course in the context of this group of participants, and the CST course added most value when delivered after the AO Basic Principles course, to build on the foundational skills already learnt.

CST has value beyond merely acquiring surgical skills. Through the environmental fidelity of the simulation, the ‘pervasive pedagogic action’ of the symbolically structured environment of the operating theatre can be recreated, enabling surgeons-in-training to gain a myriad of non-technical skills and become ‘stirred into practice’ by learning the values, behaviours and tacit knowledge required for surgical practice.

An important direction for future work in this area would be to explore the role of CST in the acquisition of non-technical skills and sociocultural capital. There is also a need to explore the experiences of the consultant trainers in delivering CST.

8.8 Reflections

In this chapters' reflections, I will consider what I have learnt in undertaking this piece of research work.

Prior to starting this PhD, I was aware of the reputation of qualitative research amongst surgeons as being a 'bit fluffy'. On doing a scoping literature search in the early weeks of the project, I could not find much qualitative surgical education research in the literature, and what little I could find was from the disciplines of sociology and anthropology. I wondered what the reason was for this deficit – was it an important but neglected area of surgical educational research, or was it simply not worth doing in the first place?

I was aware of the recent trend in clinical trials to include a qualitative component alongside the main quantitative research package, and that this was gradually becoming a requirement of major funding bodies. It seemed that qualitative research was gradually gaining recognition in T&O as being a valuable adjunct to the traditional, quantitative methods which were already familiar to me.

To find out more about it, and to help inform the decision about whether to include some qualitative work in this PhD, I undertook a qualitative research methods course early in my first year. Despite my prior prejudice, I found it fascinating. I could clearly see the benefit this would bring to the project, in providing some 'colour' to interpret the trial results.

Doing the course, and the subsequent reading for the work in this chapter, forced me to totally reframe what I had thus far taken to be absolute truths in terms of research evidence and rigor. I had spent my academic career being taught that the randomised trial is king, and everything else is inferior, and I had begun this PhD holding that bias. I found learning about the various epistemological and ontological positions really interesting, and it greatly broadened my understanding of what constitutes truth and knowledge. I came to realise that both qualitative and quantitative approaches had their respective strengths and weaknesses, and could be

complementary when applied together. In particular I came to value how qualitative research actually has better tools for investigating complex real-world phenomena such as educational experiences. I could see that in this thesis, a randomised trial was the best way to find out if training intervention A or B led to objectively better patient outcomes, and a qualitative enquiry was needed to understand how and why.

In the design phase I did a mixed-methods research course to try and better understand how these two different paradigms can be combined. In terms of mixed-methods design, what I have done in this thesis is most closely aligned to a concurrent nested mixed-methods design⁴⁸⁹, as I sought to qualitatively examine the process of the CST intervention to aid in the interpretation of the quantitative trial results. The qualitative part of the study therefore exists to add to and help explain the trial results. I chose this design because it was simple and achievable for me to do as a doctoral researcher, and seemed to be a good fit for the research question.

I do not claim to call the work in this thesis true mixed-methods research. I decided not to do any formal mixed-methods data synthesis by transformation or triangulation, and have treated the datasets independently, by presenting them in separate chapters. I did not think a mixed-methods synthesis was necessary as the results from the two chapters seem (to me at least) to be naturally complementary. I like the fact they are two separate pieces of work that look at different facets of the same research question, and did not think there was anything other than unnecessary complexity to be gained from a formal mixed-methods analysis in this case.

As well as coming to understand the value of qualitative research, the other major learning outcome for me in doing this work was in gaining an appreciation of how difficult it is to actually do it properly. What I had initially assumed would involve just a 'bit of a chat' with the study participants (in keeping with my prior prejudice about qualitative research), actually turned out to be a much more complex endeavour. I found the level of organisation and planning around the interviews to be more than I had expected. Some participants were easy to interview, and did not need much 'steering'. Others needed more encouragement and use of closed questioning. A couple of the participants are my friends in real-life and so I had to

*concentrate particularly hard on keeping to the topic guide and not allowing the conversation to wander off into a 'chat'. I also found technical factors such as recording/audio quality to be something I had not been mindful enough of to begin with, and the first interview transcript in particular came back from the typists with ****inaudible at XXh:XXm:XXs**** dotted throughout. An inescapable childcare failure of mine for one of the telephone interviews meant ****baby crying**** featured heavily in that transcript. I subsequently discovered the clinical trials unit has specialized sound-proof booths for use for telephone interviews with study participants. I would be sure to make use of these in future!*

Despite these challenges, I think the results of this chapter are really interesting, and provide valuable and previously unreported information about how cadaveric simulation training works. I enjoyed the creativity and freedom of this part of the project, and I look forward to doing more of this kind of research in the future. I can see that it will be a valuable set of tools for a career as clinical academic with a surgical education research interest.

I am pleased to have had two papers published from the qualitative work presented in this chapter.

Chapter 9: Discussion and Conclusions

In this final chapter, I will review the objectives of my thesis and summarise the new findings. I will discuss the implications of these for surgical educators and for educational theory. I will also identify the keys areas for future research in the field.

Declarations

None

9.1 Review of thesis objectives

In this thesis I set out to explore the use of cadaveric simulation for postgraduate specialist training of trauma & orthopaedic surgeons. My three objectives were;

- To make the case for using cadaveric simulation for postgraduate surgical training
- To explore how the technical skills of surgeons-in-training can be assessed
- To measure the impact of a cadaveric simulation training intervention on real world performance.

I will summarize the new findings from the work presented in this thesis to illustrate how these three objectives have been met, and situate these in the current knowledge base.

9.2 Summary of new findings and context within existing knowledge

In chapters 1 and 2, I presented a detailed, narrative overview of the present challenges in delivering surgical training, and of the emerging role of simulation in addressing these. I gave a descriptive summary of the relevant educational theory that underpins the use of simulation for surgical training.

In chapter 3, I presented a systematic review of the current evidence base for the use of cadaveric simulation for postgraduate surgical training. This is the first report on the evidence of training impact of cadaveric simulation with proper, systematic assessment of methodological rigor and level of evidence, and addresses a known gap in the literature. The results of this review showed that there is a plethora of low-quality evidence showing that cadaveric simulation is a popular training tool which may be able to induce short term objective behavioural change when measured in the simulation laboratory. In this review I identified that there was a lack of evidence of skill retention longitudinally after cadaveric simulation training, a lack of evidence

of transfer of skills into the ‘live’ operating theatre, and a lack of evidence of patient benefit following the training.

In chapter 4, I conducted the first detailed national survey of simulation provision in T&O training programmes in the UK and RoI. This was required to provide a comprehensive understanding of what was currently available in practice, and is at the forefront of the Specialist Advisory Committee current work on curriculum reform in T&O.

The results of the survey showed that there is currently widespread, but regionally variable, provision of simulation in T&O training in the UK and RoI. The availability of resources for simulation training varied widely, and provision of cadaveric simulation is likely to be influenced by the geographic and financial relationship of training programs with medical schools, where wet-laboratory facilities are generally located. The survey results also showed that the funding landscape for simulation training in T&O was complex, revealed several barriers to delivery within training programmes, and widespread reliance on funding from industry.

I then moved the focus towards how to measure technical operative skills. In chapter 5, I presented a systematic review of the current tools for assessing technical skill acquisition and operative competence in orthopaedic training. This is the first comprehensive and systematic analysis of the tools currently available for assessing technical skill and competence in T&O training. The results of the review showed that the utility evidence for the current gold standard assessment tool in T&O training in the UK, the procedure based assessment, has inadequate utility evidence to support its continued use in summative high stakes assessment of competency in training.

The review results showed that development of an assessment tool that is generalisable to the broad range of technical and non-technical skills relevant to T&O, that satisfies the utility criteria, is cost-effective, feasible and educationally impactful requires development. This review will serve, as commented by the journal reviewers, as an important benchmark for this future development work.

In knowing that I wanted to run a randomised trial of cadaveric simulation vs. standard training to look at the impact of the training on *patient outcomes*, it was clear from the results of the review in chapter 5 that ‘final product’ analysis using post-operative radiographs of real life operations was the most sensible outcome measure. This choice was mainly driven by the fact that FPA captures outcome very close to the time of surgery and is directly related to technical skill. It therefore captures a point in time before other variables beyond the surgeons’ skill and immediate peri-operative care can influence patient outcome (please see figure 21).

To address the question of what measurements should be taken (i.e. those that were feasible to measure, were representative of parameters that are influenced by technical skills, and which influence clinical outcome – figure 22) I ran a modified Delphi consensus setting exercise, which is detailed in chapter 6. The results of this work generated a core primary FPA outcomes set for the trial. These outcomes subsequently showed early evidence of face validity, construct validity and responsiveness.

Whilst I make no claims that this is a fully-formed outcome measurement tool, it is a promising start in the under-researched area of using real-world clinical outcomes for assessing technical skill and competence. There is some interesting further work to be done in exploring the concurrent validity and reliability of these measures, and their generalisability to other settings.

The main thrust of this thesis is a randomised trial comparing cadaveric simulation versus standard training for junior orthopaedic surgeons-in-training performing basic trauma operations. The trial makes a significant contribution to the evidence base as it is the first Kirkpatrick level 4 study showing patient benefit from cadaveric simulation training.

The primary research question was ‘does cadaveric simulation training lead to better patient outcomes for DHS, hemiarthroplasty and ankle fracture fixation as compared to standard training for junior postgraduate orthopaedic surgeons-in-training’.

The trial results answered this question, showing that cadaveric simulation training improved patient outcomes. There were significant observed improvements in implant positions for all the operations types, with an associated reduction in risk of failure. There was a lower acute complication rate amongst the hip fracture patients whose surgeon had undergone cadaveric simulation training. These patients generally had a quicker operation, less intra-operative blood loss, were significantly less likely to require a blood transfusion post-operatively (hemiarthroplasty) and had a shorter average inpatient stay in hospital after their operation.

We know that hip fracture is a major public health burden, and that there are about 65,000 hip fracture operations performed in the UK every year. Ankle fractures are also a common injury, with an estimated incidence of 11 per 1000 people per year⁴⁹⁰. The vast majority of the operations for both hip and ankle fractures in the NHS are performed by junior orthopaedic surgeons-in-training. The trial results, scaled to the known annual disease burden of hip and ankle fractures in the UK, suggest there could be some substantial benefits to the NHS of cadaveric simulation for junior orthopaedic surgeons-in-training. These benefits are in the form of cost-savings from less theatre time, shorter inpatient stays and fewer re-operations for failure. The benefits are also in the form of patient safety and satisfaction - better positioned implants with fewer acute complications should mean less pain and disability from surgical failure.

The secondary research questions were ‘can the early learning curve of DHS, hemiarthroplasty and ankle fracture fixation be defined’ and ‘is it feasible to use post-operative radiographs to assess technical skill’.

With regard to defining the early surgical learning curve, temporal improvement trends were seen for implant position, procedure time and image intensifier use across the three procedures for both study groups. I fitted linear trend lines to the data (for the reasons given in section 7.6.3) so my use of the term ‘curve’ here to describe learning effects is in a conceptual rather than literal sense. In demonstrating temporal improvement, the trial results shows there is a learning ‘curve’ effect which is measurable using these objective parameters, and hence answers this research question.

Regarding the question of the feasibility of using post-operative radiographs to assess technical skill, these were found to be an appropriate and feasible assessment method in the trial. The outcome measures derived from the post-operative radiographs demonstrated face validity, construct validity and responsiveness (please see section 7.6.4 for details). Further investigation of their concurrent validity, reliability and generalisability is an area for future work.

These findings should all be interpreted in the context of the limitations of the trial, the main ones being that it was a small study, and there was no low-fidelity comparator arm.

To help explain the overall trial findings, I presented an adjunctive qualitative study of how and why cadaveric simulation influences learning in chapter 8. This is the first in-depth qualitative study to investigate the mechanism of learning with cadaveric training. The results of this study show that cadaveric simulation is highly valued by surgeons-in-training, and improves confidence and preparedness for the transition to specialist registrar. It can help offset the problems with accessing conventional training opportunities in the operating theatre because of the current challenging training climate (discussed in detail in chapter 2).

This qualitative study showed that cadaveric simulation has substantial added value over low fidelity alternatives, which goes beyond the mere fact it is ‘ultra-high’ fidelity, and beyond the acquisition of technical skills. Through the ‘pervasive pedagogic’ action of the highly realistic, multidisciplinary training environment, with attendant detail to physical and psychological fidelities, learners gain a range of non-technical skills by learning the values, behaviours and hidden knowledge curriculum required to be a surgeon. In other words, learners not only learn how to do surgery, they learn how to *be surgeons*. The qualitative study showed that cadaveric simulation provides the optimum environment for intensive ‘deliberate’ practice, and in the format described in this thesis was best delivered halfway through CT2 (PGY4).

9.2.1 Implications for educational theory

I think there are two main features of the work presented in this thesis which have implications for how the classical educational theory of learning complex skills applies to surgery.

9.2.2 Revisiting the fidelity-transfer correlation

As I have discussed in chapter 2, the conventional thinking on simulator fidelity and skill transfer (the so called fidelity-transfer correlation) is that the fidelity of the simulator should match the skill level of the learner to maximise learning and cost-effectiveness. According to this theory, junior trainees (such as the participants in the trial) should be trained using low-fidelity simulation tools such as plastic bones and bench top trainers, with cadaveric simulation training reserved for the most advanced, senior trainees at the ‘refinement’/ ‘mastery’ stage of learning.

This assumption, as I hope is implicit in the results of both the trial and qualitative study, is a gross oversimplification of both the complexity of learning surgery and the true value of cadaveric simulation for junior surgeons-in-training. This classical theory is grounded in the aviation literature, and it may very well be true that a simple flight simulator for advanced beginner pilots is just as good, and more cost-effective, than a fancy, expensive, high-fidelity one.

Flying an aeroplane is, however, a different and arguably much simpler task than performing an operation. There are limited perceptual-motor skills required for flying an aeroplane, more automation, less multidisciplinary team working, and limited variation in task environment – there are only five types of commercial aircraft in operation⁴⁹¹, in contrast to the human body with its seemingly unlimited potential for variability in both anatomy and physiology.

As surgical educators we should therefore be careful not to overextend the analogy of learning to perform surgery as being like learning to fly an aeroplane. It has

perhaps been understandable in the past to extrapolate from the aviation simulation literature for the earliest stages of surgical simulation research, being as it has been until very recently a much more developed field than the equivalent in surgery.

The time has now come to develop educational theory that is specific to surgery, and grounded in evidence from surgical education research. As to the fidelity question – I believe that junior surgeons-in-training have potentially the most to gain from ultra-high fidelity cadaveric simulation, both in terms of technical and non-technical skill development, and in terms of safeguarding patients. I believe the evidence from the research presented in this thesis supports this claim, and that the classical theory on the fidelity-transfer correlation should not be applied to surgical simulation research.

9.2.3 Cadaveric simulation for optimal deliberate practice and mastery learning

I briefly discussed the concepts of deliberate practice and mastery learning in chapter 2, where ‘deliberate practice’ is targeted, repetitive performance of skills in a focused domain with specific feedback, and ‘mastery learning’ is an instructional method whereby performance is assessed against a predefined benchmark and progression to the next stage only allowed once this has been achieved.

Cadaveric simulation provides the ideal environment for deliberate practice, for the reasons described in detail in chapters 7 and 8; intensive expert guided tuition in a highly realistic environment that is without the usual time pressures and safety constraints of the operating theatre. We did not, in the CST intervention in the trial, implement any form of benchmarking assessment, apart from doing procedure based assessments for formative learning purposes only (and as already discussed PBA’s are hopeless technical skills assessment tools, but they are what we had to hand and are the current ‘gold’ standard).

I think that it would be quite straightforward to introduce a mastery learning framework into future cadaveric simulation training courses, particularly now that I

have demonstrated some early evidence for using implant position to measure technical skill. A proficiency benchmark could be set (for example TAD <25mm for DHS, with a set limit on the number of wire passes, etc.), and the trainee must perform the procedure a set number of (as yet to be determined) times to meet this standard before being deemed ‘competent’.

9.2.4 Implications for the surgical education profession

I think objective, competency based assessment in the laboratory following simulation training, before controlled release into clinical practice, is the future direction of surgical training. It seems to me to be the best way to train surgeons efficiently, to a high standard, and to safeguard patients within the current challenging surgical training climate.

I do not think simulation could, or should, ever replace ‘real-life’ surgical training in the operating theatre, and there are some experiences even the most sophisticated simulator could never properly replicate (like haemostasis, particularly brisk haemorrhage). It is, however, a necessary and valuable adjunct to a surgical training system that is in serious danger of failing to maintain end-product quality.

I also do not believe cadaveric simulation should be used in isolation to the exclusion of other simulation modalities. There have been some spectacular developments in immersive virtual reality simulators for surgery, and some recent high quality evidence that these can be effective for training⁴⁹². There have also been exciting developments in the use of tablet/smartphone based simulators. Touch Surgery™ (Touch Surgery Labs, London, UK)⁴⁹³, so called ‘serious gaming’⁴⁹⁴ is an example. These tools have the advantage of being able to be used wherever and whenever, and leverage portions of time that might not otherwise be available for training.

9.3 Conclusions

The key findings of this thesis are;

- There is a significant quantity and quality problem with surgical training in the United Kingdom
- Simulation can help mitigate this by rapidly upskilling junior trainees in a setting remote from patients
- The current provision of simulation in T&O training programmes in the UK and RoI is widespread but patchy and inconsistently funded
- There is inadequate utility evidence to support the continued use of the procedure based assessment in high-stakes assessment of competency
- Cadaveric simulation training for junior orthopaedic surgeons-in-training leads to improved patient outcomes for three common trauma procedures compared to standard training
- Early surgical learning curves can be defined for DHS, hemiarthroplasty and ankle fracture fixation
- Final product analysis using post-operative radiographs is a feasible method of assessing technical skill for these procedures
- Cadaveric simulation training provides an optimal deliberate practice environment to learn a range of technical and non-technical skills in a complete training package

In this thesis I have attempted to make a significant contribution towards answering the ‘does it work’ and ‘how does it work’ questions. The questions of ‘what type of simulation’ and ‘when’ are priority areas for further work in the field. I have set out some specific research areas for this in section 9.4 below.

9.4 Future research

The work presented in this thesis has stimulated several further research areas. I think the most important of these are;

- A health economic analysis of the cost-effectiveness of cadaveric simulation training, based on estimated failure/re-operation rates
- A further trial to compare the real-world impact of cadaveric simulation vs. high-fidelity virtual reality simulation. This can be properly powered now that we have an idea of the effect size of the training
- Continued validation work on the FPA outcome measures, with a particular focus on concurrent validity, intra/inter-observer reliability and proficiency benchmarking (i.e. define 'competent')
- Profiling of individual learning curves, to further explore the factors that might influence the level of benefit obtained from the training.

It will be a privilege to be able to take this forward in my future career as an academic clinician, to be able to apply my research training and new knowledge for the benefit of patients.

10. References

1. Delamothe T. Founding principles. *BMJ* 2008;336(7655):1216-8.
2. Worrall M. NHS at 60: advances in surgery. *The Bulletin of the Royal College of Surgeons of England* 2008;90(5):156-58.
3. Ellis H. Operations that made history. Cambridge: Cambridge University Press 1996.
4. Chikwe J, de Souza AC, Pepper JR. No time to train the surgeons. *BMJ* 2004;328(7437):418-9.
5. Darzi A, Mackay S. Assessment of surgical competence. *Quality in Health Care* 2001;10:ii64-ii69.
6. Blencowe N, Bloor J, Sheffield J, et al. From nepotism to speed-dating: selection of specialty trainees in surgery. *Bulletin of The Royal College of Surgeons of England* 2009;91:108-10.
7. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. *N Engl J Med* 2006;355(25):2664-9.
8. Dosis A, Aggarwal R, Bello F, et al. Synchronized video and motion analysis for the assessment of procedures in the operating theater. *Arch Surg* 2005;140(3):293-9.
9. Traynor O. Surgical training in an era of reduced working hours. *Surgeon* 2011;9 Suppl 1:S1-2.
10. Walter AJ. Surgical education for the twenty-first century: beyond the apprentice model. *Obstet Gynecol Clin North Am* 2006;33(2):233-6, vii.
11. Hargreaves DH, Bowditch MG, Griffin DR. On-the-job training for surgeons. 1 ed. London: The Royal Society of Medicine Press 1997.
12. Villatoro E, Patterson JE. Training and assessment. *Surgery (Oxford)* 2008;26(10):419-23.
13. Dean B, Jones L, Garfield Roberts P, et al. What is Known About the Attributes of a Successful Surgical Trainer? A Systematic Review. *Journal of Surgical Education* 2017;74(5):843-50.
14. Wolfe JH. General surgical training--improvements and problems. Vascular Advisory Committee of the Vascular Surgical Society of Great Britain and Ireland. *Annals of the Royal College of Surgeons of England* 1998;80(3 Suppl):112-16.
15. Pellegrini VDJ. Mentoring During Residency Education: A Unique Challenge for the Surgeon? *Clinical Orthopaedics and Related Research (1976-2007)* 2006;449:143-48.
16. Greenaway D. Securing the future of excellent patient care: Final report of the independent review. London, 2013. https://www.gmc-uk.org/-/media/documents/shape-of-training-final-report_pdf-53977887.pdf (accessed 05 Feb 2020).
17. Lubowitz JH, Provencher MT, Brand JC, et al. The Apprenticeship Model for Surgical Training Is Inferior. *Arthroscopy* 2015;31(10):1847-8 .
18. The Medical Act, 1950. *Br Med J* 1950;2(4674):337-8.
19. An Introduction - The Educational Landscape. Anaesthetists as Educators. London: Royal College of Anaesthetists. <https://www.rcoa.ac.uk/anaesthetists-educators-aae/educational-resources?mini=2019-06> (accessed 12 Dec 2018).
20. Department of Health. The Report of the Working Group on Specialist Medical Training. Hospital Doctors: Training for thre Future. London: Her Majesties Stationary Office 1993:1-61. <https://archive.org/details/op1279411-1001/mode/2up> (accessed 05 Feb 2020).

21. Calman K. Hospital doctors: training for the future. *BJOG: An International Journal of Obstetrics & Gynaecology* 1995;102(5):354-56.
22. Dillner L. Senior house officers: the lost tribes. *BMJ* 1993;307(6918):1549-51.
23. Donaldson L. Unfinished business: proposals for reform of the senior house officer grade - a paper for consultation. London: Department of Health 2002.
https://webarchive.nationalarchives.gov.uk/20060109162212/http://www.dh.gov.uk/PublicationsAndStatistics/Publications/PublicationsPolicyAndGuidance/PublicationsPolicyAndGuidanceArticle/fs/en?CONTENT_ID=4007842&chk=Zx93VI (accessed 05 Feb 2020).
24. House of Commons. Health - Third Report. London: The Stationary Office, 2008.
<https://publications.parliament.uk/pa/cm200708/cmselect/cmhealth/25/2505.htm> (accessed 05 Feb 2020).
25. Bonner J. The MMC story. *BMJ* 2006;333(7579):s203-s05.
26. Department of Health. Choice and opportunity: Modernising medical careers for Non-Consultant Career Grade Doctors. London: The Stationary Office, 2003.
27. Department of Health. The NHS Plan: a plan for investment, a plan for reform. London: The Stationary Office, 2000.
https://webarchive.nationalarchives.gov.uk/20121102184216/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4002960 (accessed 05 Feb 2020).
28. Morris-Stiff GJ, Sarasin S, Edwards P, et al. The European Working Time Directive: One for all and all for one? *Surgery* 2005;137(3):293-7.
29. Marron C, Byrnes C, Kirk S. An EWTD-compliant shift rota decreases training opportunities. *Bulletin of the Royal College of Surgeons of England* 2005;87(7):246-48.
30. Taffinder NJ, McManus IC, Gul Y, et al. Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. *Lancet* 1998;352(9135):1191.
31. Williamson AM, Feyer A-M. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational and Environmental Medicine* 2000;57(10):649-55.
32. Pickersgill T. The European working time directive for doctors in training. *BMJ* 2001;323(7324):1266.
33. Weinger MB, Ancoli-Israel S. Sleep deprivation and clinical performance. *JAMA* 2002;287(8):955-7.
34. Bates T, Cecil E, Greene I. The effect of the EWTD on training in general surgery: an analysis of electronic logbook records. *The Bulletin of the Royal College of Surgeons of England* 2007;89(3):106-09.
35. Moore J. Education and Training: A report from the NHS Future Forum. 2009
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213751/dh_127543.pdf (accessed 06 Feb 2020).
36. De Souza B, Nawaz S, Critchley C. Demystifying PMETB. *BMJ* 2006;332(7544):s128-s29.
37. Neville E. Modernising medical careers. *Clin Med (Lond)* 2003;3(6):529-31.
38. The Foundation Programme. <http://www.foundationprogramme.nhs.uk/> (accessed 05 Feb 2020).
39. Department of Health. The next steps - The Future Shape of Foundation, Specialist and General Practice Training Programmes. London: The Stationary Office, 2004.
<https://webarchive.nationalarchives.gov.uk/20110929193948/http://www.dh.gov.uk/>

prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4079532.pdf (accessed 05 Feb 2020).

40. Reid W. Supported from the start; ready for the future; The Postgraduate Medical Foundation Programme Review. London: Health Education England, 2019.

41. MMC: mass medical culling. *Lancet* 2007;369(9565):879.

42. Remedy UK Ltd, R (on the application of) v The General Medical Council. England and Wales High Court (Administrative Court) 28 May 2010.

43. Ribeiro B. MTAS resignations: implications for MMC. *The Bulletin of the Royal College of Surgeons of England* 2007;89(5):154-55.

44. Tooke, J. Aspiring to Excellence. Final Report of the Independent Inquiry into Modernising Medical Careers. London: Aldridge Press 2008.

http://www.asit.org/assets/documents/MMC_FINAL_REPORT_REVD_4jan.pdf (accessed 05 Feb 2020).

45. Delamothe T. Modernising Medical Careers: final report. *BMJ* 2008;336(7635):54-55.

46. Ribeiro B. The Tooke report: implications for surgery. *The Bulletin of the Royal College of Surgeons of England* 2008;90(2):42-43.

47. Royal College of Physicians: Statement regarding the government response to the Independent Review of MMC. London, 2008.

<https://www.politics.co.uk/opinion-formers/royal-college-of-physicians/article/royal-college-of-physicians-statement-regarding-the-governme> (accessed 05 Feb 2020).

48. Kmietowicz Z. Tooke report proposes a new body to get postgraduate medical training back on track. *BMJ (Clinical research ed)* 2008;336(7635):61-61.

49. Gallagher K, Dawson-Bowling S, Nawaz Z, et al. Training structures in trauma and orthopaedic surgery. *BMJ* 2010;341:c4791.

50. General Medical Council. Shape of Training: Report from the UK Shape of Training Steering Group. Edinburgh: The Scottish Government 2017.

https://www.gmc-uk.org/-/media/documents/report-from-the-uk-shape-of-training-steering-group_pdf-79105880.pdf (accessed 05 Feb 2020).

51. Limb D. The Shape of Training Report - Response of the British Orthopaedic Association (Provisional) London: BOA Education Board 2014.

52. Maruthappu M, Sood HS, Keogh B. The NHS Five Year Forward View: transforming care. *Br J Gen Pract* 2014;64(629):635-35.

53. NHS England, Public Health England, Health Education England, Monitor, Quality Care Commission. NHS Trust Development Authority Five Year Forward View. London 2014. <https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf> (accessed 05 Feb 2020).

54. Royal College of Surgeons of England. Improving Surgical Training - Trainee Prospectus. London: The Royal College of Surgeons, 2018.

55. Joint Committee on Surgical Training. <https://www.jcst.org/> (accessed 05 Feb 2020).

56. Royal College of Surgeons. Improving Surgical Training: Proposal for a pilot surgical training programme. London: RCS Professional Standards 2015.

[file:///C:/Users/mdpkfq/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempState/Downloads/improving_surgical_training_text%20\(1\).pdf](file:///C:/Users/mdpkfq/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempState/Downloads/improving_surgical_training_text%20(1).pdf) (accessed 05 Feb 2020).

57. The Association of Surgeons in Training and British Orthopaedic Trainees Association. Evidence submitted to The Royal College of Surgeons of England 'Improving Surgical Training' Consultation. London, November 2015.

http://www.asit.org/assets/documents/ASiT_and_BOTA_response_to_RCS_Improving_Surgical_Training_3.pdf (accessed 05 Feb 2020).

58. Nally D, Brown M. Improving Surgical Training: Letter To Health Education England. British Orthopaedic Trainees Association. London, June 2019. <https://www.bota.org.uk/improving-surgical-training-letter-to-health-education-england/> (accessed 05 Feb 2020).

59. NHS Digital. Hospital Admitted Patient Care Activity 2017-18: National Health Service, 2018. <https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2017-18> (accessed 05 Feb 2020).

60. Barnett K, Mercer SW, Norbury M, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012;380(9836):37-43.

61. Licchetta M, Stelmach M. Fiscal sustainability and public spending on health. London: Office for Budget Responsibility, 2016. https://obr.uk/docs/dlm_uploads/Health-FSAP.pdf (accessed 05 Feb 2020).

62. Office for National Statistics. National Population Projections: 2018-based. London, 2019. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2018based> (accessed 05 Feb 2020).

63. National Health Service. NHS Performance Statistics. London, 2019. <https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2019/12/Combined-Performance-Summary-December-October-November-data-2019-43ui4.pdf> (accessed 05 Feb 2020).

64. The Kings Fund. Trusts in deficit. 2018. <https://www.kingsfund.org.uk/projects/nhs-in-a-nutshell/trusts-deficit> (accessed 05 Feb 2020).

65. British Broadcasting Corporation. General Election 2019: Don't politicise health service - NHS boss. 2019 <https://www.bbc.co.uk/news/election-2019-50282333> (accessed 30 Jan 2020).

66. Wilford G. NHS winter crisis: more than 5,000 patients spent over 12 hours in A&E last week, new data shows. *The Telegraph* 2019. <https://www.telegraph.co.uk/news/2019/12/08/nhs-winter-crisis-5000-patients-spent-12-hours-ae-last-week/> (accessed 05 Feb 2020).

67. Coleman C. Impact of Fiscal Policies on Recruitment, Retention and Conditions of NHS Staff Debate. House of Lords. London: House of Lords Library, 2017. <https://researchbriefings.parliament.uk/ResearchBriefing/Summary/LLN-2017-0087> (accessed 05 Feb 2020).

68. General Medical Council. The state of medical education and practice in the UK. London, 2017. <https://www.gmc-uk.org/static/documents/content/SoMEP-2017-final-executive-summary.pdf> (accessed 05 Feb 2020).

69. Rimmer, A. HEE staff cuts threaten medical education and patient care, BMA and unions warn. *BMJ* 2017; 356:j1069.

70. Eardley, I. A question of balance: The extended surgical team. London: The Royal College of Surgeons of England, 2016. [file:///C:/Users/mdpkfq/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempState/Downloads/est_2016_web2%20\(1\).pdf](file:///C:/Users/mdpkfq/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempState/Downloads/est_2016_web2%20(1).pdf) (accessed 05 Feb 2020).

71. Francis R. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry. Stationary Office, London: The Stationary Office, 2013.
<https://www.gov.uk/government/publications/report-of-the-mid-staffordshire-nhs-foundation-trust-public-inquiry> (accessed 05 Feb 2020).
72. Berwick D. A promise to learn - a commitment to act: improving the safety of patients in England. London, Williams Lea. 2013.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/226703/Berwick_Report.pdf (accessed 05 Feb 2020).
73. Department of Health. Culture change in the NHS. Applying the lessons of the Francis Inquiries. Stationary Office. London: Williams Lea Group on behalf of The Stationary Office, 2015.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/403010/culture-change-nhs.pdf (accessed 05 Feb 2020).
74. Francis R. Freedom to speak up. An independent review into creating an open and honest reporting culture in the NHS. London, 2015.
https://webarchive.nationalarchives.gov.uk/20150218150953/https://freedomtospeakup.org.uk/wp-content/uploads/2014/07/F2SU_web.pdf (accessed 05 Feb 2020).
75. Haynes AB, Weiser TG, Berry WR, et al. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine* 2009;360(5):491-99.
76. Michaels R, Makary MA, Dahab Y, Frassica FJ. Achieving the National Quality Forum’s “Never Events”: Prevention of Wrong Site, Wrong Procedure, and Wrong Patient Operations. *Annals of Surgery* 2007;245(4):526-32.
77. Consultant Outcomes Data . <https://www.england.nhs.uk/2014/11/outcome-publication/> (accessed 05 Feb 2020).
78. National Joint Registry Mission Statement 2018.
<http://www.njrcentre.org.uk/njrcentre/AbouttheNJR/tabid/73/Default.aspx>. (accessed 05 Feb 2020).
79. House of Commons Public Administration Select Committee. Investigating clinical incidents in the NHS. Sixth Report of Session 2014-15. London: The Stationary Office, 2015.
<https://publications.parliament.uk/pa/cm201415/cmselect/cmpubadm/886/88601.htm> (accessed 05 Feb 2020).
80. Gaskell N, Hinton R, Page T, et al. Putting an end to Black Wednesday: improving patient safety by achieving comprehensive trust induction and mandatory training by day 1 *Clinical Medicine* 2016;16(2):124-28.
81. Hayes L. Improving junior doctor handover between jobs. *BMJ Quality Improvement Reports* 2014;3(1).
82. Baldwin PJ, Dodd M, Wrate RM. Junior doctors making mistakes. *The Lancet* 1998;351(9105):804.
83. Aggarwal R, Darzi A. Technical-Skills Training in the 21st Century. *New England Journal of Medicine* 2006;355(25):2695-96.
84. Curran I. A Framework for Technology Enhanced Learning. London: Department of Health, 2011.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215316/dh_131061.pdf (accessed 05 Feb 2020).
85. House J. Calling time on doctors’ working hours. *Lancet* 2009;373(9680):2011-12.

86. Temple J. Time for Training. A Review of the impact of the European Working Time Directive on the quality of training. London: Medical Education England, 2010.
87. Rashid A, Al-Hadithy N, Rossouw D, et al. Consultant Attitudes to 'St' Higher Surgical Trainees. *The Bulletin of the Royal College of Surgeons of England* 2012;94(10):1-334.
88. Purcell Jackson G, Tarpley JL. How long does it take to train a surgeon? *BMJ* 2009;339.
89. Mitchell J, Hayhurst C, Robinson SM. Can a senior house officer's time be used more effectively? *Emergency Medicine Journal* 2004;21(5):545-47.
90. Rooney A, Moideen AN, Akinyooye A, et al. An audit of clinical training exposure amongst junior doctors working in Trauma & Orthopaedic Surgery in 101 hospitals in the United Kingdom. *BMC Medical Education* 2018;18(1):1.
91. Jameson SS, Gupta S, Lamb A, et al. Operative Experience in Trauma and Orthopaedics: Effect of Changes to Trainees' Working Patterns. *The Bulletin of the Royal College of Surgeons of England* 2010;92(9):1-5.
92. Hopkins ME, Harris AS, Davids JD, et al. Are core surgical trainees being trained to operate? *The Bulletin of the Royal College of Surgeons of England* 2019;101(3):106-11.
93. Toll EC, Davis CR. More Trainees and Less Operative Exposure: A Quantitative Analysis of Training Opportunities for Junior Surgical Trainees. *The Bulletin of the Royal College of Surgeons of England* 2010;92(5):170-73.
94. James H, Fisher JG, Griffin DR, et al. If we can't get to theatre, we can't learn to operate: A qualitative study of factors influencing core trainee access to the operating theatre in Trauma & Orthopaedics. *Ann R Coll Surg Engl* 2019; In Press.
95. General Medical Council. National Training Survey. London, 2018. https://www.gmc-uk.org/-/media/documents/dc11391-nts-2018-initial-findings-report_pdf-75268532.pdf (accessed 05 Feb 2020).
96. General Medical Council. Training Environments 2017: Key findings from the National Training Surveys. London, 2017. https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/-/media/documents/nts-report-20173_pdf-72689147.pdf (accessed 05 Feb 2020).
97. General Medical Council. The State of Medical Education and Practice in the UK. London, 2016. https://www.gmc-uk.org/-/media/documents/GMC_SOMEPEP_2016___Executive_summary.pdf_68137466.pdf (accessed 05 Feb 2020).
98. Gilbert A, Hockey P, Vaithianathan R, et al. Perceptions of junior doctors in the NHS about their training: results of a regional questionnaire. *BMJ Quality & Safety Online* 2012:1-5.
99. O'Callaghan J, Mohan HM, Sharrock A, et al. Cross-sectional study of the financial cost of training to the surgical trainee in the UK and Ireland. *BMJ Open* 2017;7(11):e018086.
100. Moak TN, Cress PE, Tenenbaum M, et al. The Leaky Pipeline of Women in Plastic Surgery: Embracing Diversity to Close the Gender Disparity Gap. *Aesthetic Surgery Journal* 2019.
101. Khoushhal Z, Hussain MA, Greco E, et al. Prevalence and Causes of Attrition Among Surgical Residents: A Systematic Review and Meta-analysis. *JAMA Surgery* 2017;152(3):265-72.
102. Cochran A, Hauschild T, Elder WB, et al. Perceived gender-based barriers to careers in academic surgery. *The American Journal of Surgery* 2013;206(2):263-68.

103. Chen MM, Yeo HL, Roman SA, et al. Life events during surgical residency have different effects on women and men over time. *Surgery* 2013;154(2):162-70.
104. Criado-Perez C. Invisible women: data bias in a world designed for men. New York: Abrams Press, 2019.
105. Smith R. Hospital trusts ripping off junior doctors by charging for accomodation. *The Telegraph* 2008. <https://www.telegraph.co.uk/news/2268410/Hospital-trusts-ripping-off-junior-doctors-by-charging-for-accommodation.html> (accessed 05 Feb 2020).
106. Batt-Rawden S. If we don't start looking after our doctors there will be no one to care for our patients. *Metro* 2020. <https://metro.co.uk/2020/01/13/dont-look-doctors-no-one-care-patients-12009732/> (accessed 05 Feb 2020).
107. McClelland L, Holland J, Lomas JP, et al. A national survey of the effects of fatigue on trainees in anaesthesia in the UK. *Anaesthesia* 2017;72(9):1069-77.
108. Rimmer A. The future of doctors' messes. *BMJ* 2019;364:k5367.
109. Campbell D. Dossier reveals 'petty tortures' of NHS trainee doctors denied leave. *The Guardian* 2019. <https://www.theguardian.com/society/2019/mar/28/nhs-trainee-doctors-denied-leave-dossier-hospitals> (accessed 05 Feb 2020).
110. Wilkinson E. How to fix doctors' rotas. *BMJ* 2019;365:l4367.
111. Kay A. This is going to hurt: secret diaries of a junior doctor. London: Picador 2017.
112. Lipworth W, Little M, Markham P, et al. Doctors on Status and Respect: A Qualitative Study. *Journal of Bioethical Inquiry* 2013;10(2):205-17.
113. Timmermans S, Oh H. The continued social transformation of the medical profession. *Journal of Health and Social Behaviour* 2010;51(S):S94-S106.
114. Pearson K. Taking revalidation forward: improving the process of relicensing for doctors. London: General Medical Council, 2017.
115. Lakhani M. A way forward. *BMJ (Clinical research ed)* 2005;330(7503):1326-28
116. Massey C. The Act needs action: GMC Wordpress; 2017. <https://gmcuk.wordpress.com/2017/07/17/the-act-needs-action/> (accessed 3rd February 2020).
117. Bawa-Gaba v General Medical Council. Courts and Tribunals Judiciary. C1/2018/0356, 2018.
118. Vaughan J. Post Bawa-Garba: how do we detoxify the climate of fear? *The Bulletin of the Royal College of Surgeons of England* 2018;100(8):325.
119. Dyer C. Bawa-Garba case has left profession shaken and stirred. *BMJ* 2018;360:k456.
120. Wickware C. #IamHadiza: The full story of the legal case that shook medicine. *Pulse* 2018 <http://www.pulsetoday.co.uk/hot-topics/bawa-garba/iamhadiza-the-full-story-of-the-legal-case-that-shook-medicine/20036250.article> (accessed 06 Feb 2020).
121. Cohen D. Support mounts for paediatrician whom GMC wants struck off. *BMJ* 2017;359:j5384.
122. Jha S. To err is homicide in Britain - The case of Dr Hadiza Bawa-Garba. The Health Care Blog, 2018. <https://thehealthcareblog.com/blog/2018/01/30/to-err-is-homicide-in-britain-the-case-of-dr-hadiza-bawa-garba/> (accessed 06 Feb 2020).
123. Rimmer A. Bawa-Garba case: GMC chief executive should resign, says consultants' union. *BMJ* 2018;362:k3673.
124. Spooner S, Gibson J, Rigby D, et al. Stick or twist? Career decision-making during contractual uncertainty for NHS junior doctors. *BMJ Open* 2017;7(1).

125. Brecknell J. Recruitment & retention in surgery. London: Health Education England, 2018. file:///C:/Users/mdpkfq/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempState/Downloads/300pm%20Recruitment%20and%20retention%20John%20Brecknell%20(1).pdf (accessed 06 Feb 2020).
126. Rimmer A. Surgeons are less satisfied with training than other doctors, GMC survey finds. *BMJ* 2014;348:g4330.
127. Junior doctor morale: Understanding best practice working environments. London: Health Education England, 2017. https://www.hee.nhs.uk/sites/default/files/documents/Junior%20Doctors'%20Morale%20-%20understanding%20best%20practice%20working%20environments_0.pdf (accessed 06 Feb 2020).
128. Goddard AF. Lessons to be learned from the UK Junior Doctors' Strike. *JAMA* 2016;316(14):1445-46.
129. Women In Surgery Statistics, Royal College of Surgeons, 2020. <https://www.rcseng.ac.uk/careers-in-surgery/women-in-surgery/statistics/>. (accessed 06 Feb 2020).
130. British Orthopaedic Trainees Association - Women In Surgery <http://www.bota.org.uk/women-in-surgery/> (accessed 06 Feb 2020).
131. Clinical Academic Careers Framework: A framework for optimising clinical academic careers across healthcare professions. London: Health Education England, 2014. https://www.hee.nhs.uk/sites/default/files/documents/HEE_Clinical_Academic_Careers_Framework.pdf (accessed Feb 06 2020).
132. Clinical Academic Careers Framework: A framework for optimising clinical academic careers across healthcare professions, revised edition. London: Health Education England, 2018. <https://www.hee.nhs.uk/sites/default/files/documents/2018-02%20CAC%20Framework.pdf> (accessed 06 Feb 2020).
133. Panagioti M, Khan K, Keers RN, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. *BMJ* 2019;366:l4185.
134. Society for Simulation in Healthcare. <https://www.ssih.org/About-SSH/About-Simulation> (accessed 06 Feb 2020).
135. Aggarwal R, Mytton OT, Derbrew M, et al. Training and simulation for patient safety. *Quality and Safety in Health Care* 2010;19(Suppl 2):i34.
136. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. *BJS (British Journal of Surgery)* 2004;91(12):1549-58.
137. Reznick R. Testing technical skills outside the operating room: an innovative bench model examination. *American Journal of Surgery* 1997;173:226-30.
138. Beaubien JM, Baker DP. The use of simulation for training teamwork skills in health care: how low can you go? *Quality and Safety in Health Care* 2004;13(suppl 1):i51.
139. Marshall SD, Flanagan B. Simulation-based education for building clinical teams. *J Emerg Trauma Shock* 2010;3(4):360-68.
140. Khudyakov A, Jean C, Jankovic M. Simulation Methods in the Healthcare Systems. In: Aiguier M, Boulanger F, Krob D, et al., eds. *Complex Systems Design & Management*: Springer, Cham 2014.

141. Carter FJ, Schijven MP, Aggarwal R, et al. Consensus guidelines for validation of virtual reality surgical simulators. *Surgical Endoscopy And Other Interventional Techniques* 2005;19(12):1523-32.
142. Mitchell E, Arora S. How educational theory and practice can inform the training and practice of vascular surgeons. *Journal of Vascular Surgery* 2012;56(2):530-37.
143. White C, Rodger W, Tang T. Current understanding of learning psychomotor skills and the impact on teaching laparoscopic surgical skills. *The Obstetrician & Gynaecologist* 2016;18:53-63.
144. Shah A, Barto A. Effect on movement selection of an evolving sensory representation: A multiple controller model of skill acquisition. *Brain Research* 2009;3:55-73.
145. Ungerleider L, Doyon J, Karni A. Imaging Brain Plasticity during Motor Skill Learning. *Neurobiology of Learning and Memory* 2002;78(3):553-64.
146. Fitts P, Posner M. Human Performance. Belmont, CA: Brooks Cole 1967.
147. Dreyfus S, Dreyfus H. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. University of California, 1980.
148. Carraccio C, Benson B, Nixon L, et al. From the Educational Bench to the Clinical Bedside: Translating the Dreyfus Developmental Model to the Learning of Clinical Skills. *Academic Medicine* 2008;83(3):761-67.
149. Brydges R, Carnahan H, Backstein D, et al. Application of Motor Learning Principles to Complex Surgical Tasks: Searching for the Optimal Practice Schedule. *Journal of Motor Behaviour* 2007;39(1):40-48.
150. Naylor JC, Briggs GE. Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology* 1963;65(3):217-24.
151. Hansen S, Tremblay L, Elliot D. Part and whole practice: chunking and online control in the acquisition of a serial motor task. *Res Q Exerc Sport* 2005;76(1):60-6.
152. Park J, Wilde H, Shea C. Part-whole practice movement of sequences. *J Mot Behav* 2004;36(1):51-61.
153. Sattelmayer M, Elsig S, Hilfiker R, et al. A systematic review and meta-analysis of selected motor learning principles in physiotherapy and medical education. *BMC Medical Education* 2016;16(15):1-22.
154. Ericsson K. Deliberate Practice and Acquisition of Expert Performance: A General Overview. *Acad Emerg Med* 2008;15(11):988-94.
155. Ericsson K, Krampe T, Tesch-Romer C. The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review* 1993;100(3):363-406.
156. Bieliauskas LA, Langenecker S, Graver C, et al. Cognitive Changes and Retirement among Senior Surgeons (CCRASS): Results from the CCRASS Study. *Journal of the American College of Surgeons* 2008;207(1):69-78.
157. Blasier RB. The Problem of the Aging Surgeon: When Surgeon Age Becomes a Surgical Risk Factor. *Clinical Orthopaedics and Related Research* 2009;467(2):402-11.
158. Waljee JF, Greenfield LJ. Aging and Surgeon Performance. *Advances in Surgery* 2007;41:189-98.
159. MacCaskie A, Kenny D, Deshmukh S. How can surgical training benefit from theories of skilled motor development, musical skill acquisition and performance psychology? *The Medical Journal of Australia* 2011;194(9):463-465.
160. Mazzon G, Sridhar A, Busuttill G, et al. Learning Curves for Robotic Surgery: a Review of the Recent Literature. *Current Urology Reports* 2017;18(11):89.

161. Simpson AHRW, Howie CR, Norrie J. Surgical trial design – learning curve and surgeon volume. *Bone & Joint Research* 2017;6(4):194-95.
162. Department of Health. 150 Years of the Annual Report of the Chief Medical Officer. London: APS Group, 2009.
163. Department of Health. A high quality workforce: NHS Next Stage Review. London: The Stationary Office 2008.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf (accessed 06 Feb 2020)
164. General Medical Council. The Trainee Doctor. London, 2011.
http://fpmalta.com/uploads/2013/GMC%20Trainee_Doctor%202011.pdf (accessed 06 Feb 2020)
165. Sarker S, Patel B. Simulation and surgical training. *International Journal of Clinical Practice* 2007;61(12):2120-5.
166. Munz Y, Kumar B, Moorthy K, et al. Laparoscopic virtual reality and box trainers. *Surg Endosc* 2004;18:485-94.
167. Atesok K, Mabrey J, Jazrawi L, et al. Surgical Simulation in Orthopaedic Skills Training. *J Am Acad Surg* 2012;20:410-22.
168. Kneebone R. Simulation in surgical training: educational issues and practical implications. *Medical Education* 2003;37(3):267-77.
169. Chmarra M, Klein S, de Winter J, et al. Objective classification of residents based on their psychomotor laparoscopic skills. *Surgical Endoscopy* 2010;24(5):1031-39.
170. Debes A, Aggarwal R, Balasundaram I, et al. Construction of an evidence-based, graduated training curriculum for D-box, a webcam-based laparoscopic basic skills trainer box. *American Journal of Surgery* 2012;203(6):768-75.
171. Gallagher A, Richie K, McClure N, et al. Objective psychomotor skills assessment of experienced, junior and novice laparoscopists with virtual reality. *World Journal of Surgery* 2001;25(11):1478-83.
172. Gor M, McCloy R, Stone R, et al. Virtual reality laparoscopic simulator for assessment in gynaecology. *BJOG* 2003;110(2):181-87.
173. Grober E, Roberts M, Shin E, et al. Intraoperative assessment of technical skills on live patients using economy of hand motion: establishing learning curves of surgical competence. *American Journal of Surgery* 2010;199(1):81-85.
174. Sturm P, Windsor J, Cosman P, et al. Systematic Review of Skills Transfer after Surgical Simulation Training. *Annals of Surgery* 2008;248:166-79.
175. Giger U, Fresard I, Hafliger A, et al. Laparoscopic training on Thiel human cadavers: A model to teach advanced laparoscopic procedures. *Surgical Endoscopy* 2008;22(4):901-06.
176. Gilbody J, Prasthofer A, Ho K, et al. The use and effectiveness of cadaveric workshops in higher surgical training: a systematic review. *Annals of the Royal College of Surgeons of England* 2011;93(5):347-56.
177. Stirling ERB, Lewis TL, Ferran NA. Surgical skills simulation in trauma and orthopaedic training. *J Orthop Surg Res* 2014;9:126-26.
178. Sutherland L, Middleton P, Anthony A, et al. Surgical Simulation - A Systematic Review. *Annals of Surgery* 2006;243:291-300.
179. Da Luz L, Nascimento B, Tien C, et al. Current use of live tissue training in trauma: a descriptive systematic review. *Canadian Journal of Surgery* 2015;58(3 Suppl 3):S125-S34.
180. Human Tissue Act, 2004.
<https://www.legislation.gov.uk/ukpga/2004/30/contents> (accessed 06 Feb 2020)

181. Human Tissue Authority. <http://www.hta.gov.uk/licensingandinspections.cfm>. (accessed Feb 06 2020)
182. Carey J, Rommer E, Sheckter C, et al. Simulation of plastic surgery and microvascular procedures using perfused fresh human cadavers. *Journal of Plastic, Reconstructive and Aesthetic Surgery* 2014;67:42-48.
183. Sugand K, Abrahams P, Khurana A. The anatomy of anatomy: a review for its modernisation. *Anatomical Sciences Education* 2010;3(2):83-93.
184. Pham M, Kale A, Marquez Y, et al. A perfusion-based Human Cadaveric Model for Management of Carotid Artery Injury during Endoscopic Endonasal Skull Base Surgery. *J Neurol Surg B Skull Base* 2014;75(5):309-13.
185. Carey J, Minneti M, Leland H, et al. Perfused fresh cadavers: method for application to surgical simulation. *American Journal of Surgery* 2015;210(1):179-87.
186. Reihisen T, Alberti L, Speich J, et al. Feasibility of a perfused and ventilated cadaveric model for assessment of lifesaving traumatic haemorrhage and airway management skills. *Journal of Trauma and Acute Care Surgery* 2016;80(5):799-804.
187. Blaschko S, Brooks M, Dhuy M, et al. Co-ordinated Multiple Cadaver Use for Minimally Invasive Surgical Training. *Journal of the Society of Laparoendoscopic Surgeons* 2007;11:403-07.
188. Issenburg B, MacGahie W, Petrusa E, et al. Features and uses of high-fidelity medical simulators that lead to effective learning: a BEME systematic review. *Medical Teacher* 2005;27(1):10-28.
189. Meller G. A typology of simulators for medical education. *Journal of Digital Imaging* 1997;10(1):194-96.
190. Miller G. The Assessment of Clinical Skills/Competence/Performance. *Academic Medicine* 1990;65(9):63-67.
191. Kirkpatrick D. Techniques for evaluating training programmes. *Journal of American Society for Training and Development* 1959;13(11):11-12.
192. Buckley CE, Kavanagh DO, Traynor O, et al. Is the skillset obtained in surgical simulation transferable to the operating theatre? *The American Journal of Surgery* 2014;207(1):146-57. doi: 10.1016/j.amjsurg.2013.06.017
193. Dawe SR, Pena GN, Windsor JA, et al. Systematic review of skills transfer after surgical simulation-based training. *BJS* 2014;101(9):1063-76.
194. Palter VN, Grantcharov T, Harvey A, et al. Ex Vivo Technical Skills Training Transfers to the Operating Room and Enhances Cognitive Learning: A Randomized Controlled Trial. *Annals of Surgery* 2011;253(5):886-89.
195. Roach M. *Stiff: The Curious Lives of Human Cadavers*. London: Penguin Books 2003.
196. Ghosh S. Human cadaveric dissection: a historical account from ancient Greece to the modern era. *Anat Cell Biol* 2015;48(3):153-69.
197. Hildebrandt S. Capital punishment and anatomy: history and ethics of an ongoing association. *Clinical Anatomy* 2008;21(1):5-14.
198. Mahadevan V. The Human Tissue Act 2004: Implications for anatomical work at the college. *Ann R Coll Surg Engl (Suppl)* 2006;88:264-65.
199. Duguid A, Arundell A, Bainbridge LC. How to set up and run a cadaveric surgical simulation programme. *The Bulletin of the Royal College of Surgeons of England* 2016;98(8):358-61.
200. Cohen N. Is it just semantics? Medical students and their 'first patients'. *Journal of Medical Ethics* 2019;45(6):411.

201. Champney TH, Hildebrandt S, Gareth Jones D, et al. BODIES R US: Ethical Views on the Commercialization of the Dead in Medical Education and Research. *Anatomical Sciences Education* 2019;12(3):317-25.
202. Champney TH. A proposal for a policy on the ethical care and use of cadavers and their tissues. *Anatomical sciences education* 2011;4(1):49-52.
203. Champney TH. The business of bodies: Ethical perspectives on for-profit body donation companies. *Clinical anatomy (New York, NY)* 2016;29(1):25-9.
204. Doozandeh, P, Ritter, FE. Does Simulation Fidelity Affect Training? A lesson from a brief review of literature. International Conference on Social Computing, Behavioural-Cultural Modeling, Prediction and Behaviour Representation in Modeling and Simulation; 2019; Washington DC, USA. http://sbp-brims.org/2019/proceedings/papers/working_papers/Doozandeh_paper_50.pdf (accessed 06 Feb 2020)
205. Badash I, Burt K, Solorzano CA, et al. Innovations in surgery simulation: a review of past, current and future techniques. *Annals of translational medicine* 2016;4(23):453-53.
206. Miller RB. Psychological considerations in the design of training equipment. American Institute for Research: Aero Medical Laboratory, 1954.
207. Alessi SM. Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction* 1988;15(2):40-47.
208. Adams AJ, Wasson EA, Admire JR, et al. A Comparison of Teaching Modalities and Fidelity of Simulation Levels in Teaching Resuscitation Scenarios. *Journal of surgical education* 2015;72(5):778-85.
209. Cortegiani A, Russotto V, Montalto F, et al. Effect of High-Fidelity Simulation on Medical Students' Knowledge about Advanced Life Support: A Randomized Study. *PLOS ONE* 2015;10(5):e0125685.
210. Veaudor M, Gérinière L, Souquet P-J, et al. High-fidelity simulation self-training enables novice bronchoscopists to acquire basic bronchoscopy skills comparable to their moderately and highly experienced counterparts. *BMC Medical Education* 2018;18(1):191.
211. Tan SC, Marlow N, Field J, et al. A randomized crossover trial examining low-versus high-fidelity simulation in basic laparoscopic skills training. *Surgical Endoscopy* 2012;26(11):3207-14.
212. Hughes JL, Katsogridakis E, Junaid-Siddiqi A, et al. The development of cadaveric simulation for core surgical trainees. *The Bulletin of the Royal College of Surgeons of England* 2018;101(1):38-43.
213. Moher D, Liberati A, Tetzlaff J, et al. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine* 2009;6(7):e1000097.
214. Liberati A, Altman DG, Tetzlaff J, et al. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *Annals of Internal Medicine* 2009;151(4):W-65-W-94.
215. PROSPERO International prospective register of systematic reviews. <https://www.crd.york.ac.uk/prospero/> (accessed 06 Feb 2020)
216. Morgan M, Aydin A, Salih A, et al. Current Status of Simulation-based Training Tools in Orthopedic Surgery: A Systematic Review. *Journal of Surgical Education* 2017;74(4):698-716.

217. Cook DA, Reed DA. Appraising the Quality of Medical Education Research Methods: The Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale-Education. *Academic Medicine* 2015;90(8):1067-76.
218. Smith RP, Learman LA. A Plea for MERSQI: The Medical Education Research Study Quality Instrument. *Obstetrics & Gynecology* 2017;130(4):686-90.
219. Reed DA, Beckman TJ, Wright SM, et al. Predictive Validity Evidence for Medical Education Research Study Quality Instrument Scores: Quality of Submissions to JGIM's Medical Education Special Issue. *Journal of General Internal Medicine* 2008;23(7):903-07.
220. Lorenz R, Stechemesser B, Reinhold W, et al. Development of a standardized curriculum concept for continuing training in hernia surgery: German Hernia School. *Hernia* 2017;21(2):153-62.
221. Gunst M, O'Keeffe T, Hollett L, et al. Trauma operative skills in the era of nonoperative management: the trauma exposure course (TEC). *Journal of Trauma-Injury Infection & Critical Care* 2009;67(5):1091-6.
222. Sharma M, Horgan A. Comparison of fresh-frozen cadaver and high-fidelity virtual reality simulator as methods of laparoscopic training. *World Journal of Surgery* 2012;36(8):1732-7.
223. Gasco J, Holbrook TJ, Patel A, et al. Neurosurgery simulation in residency training: feasibility, cost, and educational benefit. *Neurosurgery* 2013;73 Suppl 1:39-45.
224. Dunnington GL. A model for teaching sentinel lymph node mapping and excision and axillary lymph node dissection. *J Am Coll Surg* 2003;197(1):119-21.
225. Supe A, Dalvi A, Prabhu R, et al. Cadaver as a model for laparoscopic training. *Indian Journal of Gastroenterology* 2005;24(3):111-3.
226. Reed AB, Crafton C, Giglia JS, et al. Back to basics: use of fresh cadavers in vascular surgery training. *Surgery* 2009;146(4):757-62; discussion 62-3.
227. Lewis CE, Peacock WJ, Tillou A, et al. A novel cadaver-based educational program in general surgery training. *Journal of Surgical Education* 2012;69(6):693-8.
228. Shekter CC, Kane JT, Minneti M, et al. Incorporation of Fresh Tissue Surgical Simulation into Plastic Surgery Education: Maximizing Extraclinical Surgical Experience. *Journal of Surgical Education* 2013;70(4):466-74.
229. Jansen S, Cowie M, Linehan J, et al. Fresh frozen cadaver workshops for advanced vascular surgical training. *ANZ Journal of Surgery* 2013;84(11):877-80.
230. Aboud E, Aboud G, Al-Mefty O, et al. "Live cadavers" for training in the management of intraoperative aneurysmal rupture. *J Neurosurg* 2015;123:1339-1346.
231. Ahmed K, Aydin A, Dasgupta P, et al. A novel cadaveric simulation program in urology. *Journal of Surgical Education* 2015;72(4):556-65.
232. Aydin A, Ahmed K, Khan MS, et al. The role of human cadaveric procedural simulation in urology training. *Journal of Urology* 2015;193(4).
233. Liu JK, Kshetry VR, Recinos PF, et al. Establishing a surgical skills laboratory and dissection curriculum for neurosurgical residency training. *Journal of Neurosurgery* 2015;123(5):1331-8.
234. Winer JL, Kramer DR, Robison RA, et al. Cerebrospinal fluid reconstitution via a perfusion-based cadaveric model: feasibility study demonstrating surgical simulation of neuroendoscopic procedures. *Journal of Neurosurgery* 2015;123(5):1316-21.

235. Clark MA, Yoo SS, Kundu RV. Nail Surgery Techniques: A Single Center Survey Study on the Effect of a Cadaveric Hand Practicum in Dermatology Resident Education. *Dermatologic Surgery* 2016;42(5):696-8.
236. Hanu-Cernat L, Thanabalan N, Alali S, et al. Effectiveness of cadaveric surgical simulation as perceived by participating OMFS trainees. *British Journal of Oral and Maxillofacial Surgery* 2016;54 (10):e126.
237. Calio BP, Kepler CK, Koerner JD, et al. Outcome of a Resident Spine Surgical Skills Training Program. *Clinical Spine Surgery : A Spine Publication* 2017;30(8):E1126-E29.
238. Nobuoka D, Yagi T, Kondo Y, et al. The utility of cadaver-based surgical training in hepatobiliary and pancreatic surgery. *Journal of Hepato-Biliary-Pancreatic Sciences* 2017;24 (Supplement 1):A338.
239. Pacca P, Jhavar SS, Seclen DV, et al. "Live Cadaver" Model for Internal Carotid Artery Injury Simulation in Endoscopic Endonasal Skull Base Surgery. *Operative neurosurgery* 2017;13(6):732-38.
240. Takayesu JK, Peak D, Stearns D. Cadaver-based training is superior to simulation training for cricothyrotomy and tube thoracostomy. *Internal & Emergency Medicine* 2017;12(1):99-102.
241. Weber EL, Leland HA, Azadgoli B, et al. Preoperative surgical rehearsal using cadaveric fresh tissue surgical simulation increases resident operative confidence. *Annals of Translational Medicine* 2017;5(15)302.
242. Bertolo R, Garisto J, Dagenais J, et al. Single Session of Robotic Human Cadaver Training: The Immediate Impact on Urology Residents in a Teaching Hospital. *Journal of Laparoendoscopic & Advanced Surgical Techniques* 2018;28(10):1157-62.
243. Chouari TAM, Lindsay K, Bradshaw E, et al. An enhanced fresh cadaveric model for reconstructive microsurgery training. *European Journal of Plastic Surgery* 2018;41(4):439-46.
244. Wynn GR, Austin RCT, Motson RW. Using cadaveric simulation to introduce the concept and skills required to start performing transanal total mesorectal excision. *Colorectal Disease* 2018;20(6):496-501.
245. AlJamal Y, Buckarma E, Ruparel R, et al. Cadaveric Dissection vs Homemade Model: What is the Best Way to Teach Endoscopic Totally Extraperitoneal Inguinal Hernia Repair? *Journal of Surgical Education* 2018;75(3):787-91.
246. Sharma G, Aycart MA, Najjar PA, et al. A cadaveric procedural anatomy course enhances operative competence. *Journal of Surgical Research* 2016;201(1):22-8.
247. Mitchell EL, Sevdalis N, Arora S, et al. A fresh cadaver laboratory to conceptualize troublesome anatomic relationships in vascular surgery. *Journal of Vascular Surgery* 2011;55(4):1187-94.
248. Robinson WP, Doucet DR, Simons JP, et al. An intensive vascular surgical skills and simulation course for vascular trainees improves procedural knowledge and self-rated procedural competence. *Journal of Vascular Surgery* 2017;65(3):907-15.e3.
249. Hazan E, Torbeck R, Connolly D, et al. Cadaveric simulation for improving surgical training in dermatology. *Dermatology online journal* 2018;24(6).
250. Sharma M, Macafee D, Horgan AF. Basic laparoscopic skills training using fresh frozen cadaver: a randomized controlled trial. *American Journal of Surgery* 2013;206(1):23-31.

251. Sundar SJ, Healy AT, Kshetry VR, et al. A pilot study of the utility of a laboratory-based spinal fixation training program for neurosurgical residents. *Journal of Neurosurgery Spine* 2016;24(5):850-6.
252. Chong W, Downing K, Leegant A, et al. Resident Knowledge, Surgical Skill, and Confidence in Transobturator Vaginal Tape Placement: The Value of a Cadaver Laboratory. *Female Pelvic Medicine & Reconstructive Surgery* 2017;23(6):392-400.
253. Anastakis DJ, Regehr G, Reznick RK, et al. Assessment of technical skills transfer from the bench training model to the human model. *American Journal of Surgery* 1999;177(2):167-70.
254. Sidhu RS, Park J, Brydges R, et al. Laboratory-based vascular anastomosis training: a randomized controlled trial evaluating the effects of bench model fidelity and level of training on skill acquisition. *Journal of Vascular Surgery* 2007;45(2):343-9.
255. Gottschalk MB, Yoon ST, Park DK, et al. Surgical training using three-dimensional simulation in placement of cervical lateral mass screws: a blinded randomized control trial. *Spine Journal: Official Journal of the North American Spine Society* 2015;15(1):168-75.
256. Camp CL, Krych AJ, Stuart MJ, et al. Improving resident performance in knee arthroscopy: A prospective value assessment of simulators and cadaveric skills laboratories. *Journal of Bone and Joint Surgery - American Volume* 2016;98(3):220-25.
257. LeBlanc F, Champagne BJ, Augestad KM, et al. A comparison of human cadaver and augmented reality simulator models for straight laparoscopic colorectal skills acquisition training. *Journal of the American College of Surgeons* 2010;211(2):250-5.
258. Zirkle M, Taplin MA, Anthony R, et al. Objective assessment of temporal bone drilling skills. *Annals of Otolaryngology, Rhinology & Laryngology* 2007;116(11):793-8.
259. Mackenzie CF, Garofalo E, Puche A, et al. Performance of Vascular Exposure and Fasciotomy Among Surgical Residents Before and After Training Compared With Experts. *JAMA Surgery* 2017;152(6):1-8.
260. Mednick Z, Tabanfar R, Alexander A, et al. Creation and validation of a simulator for corneal rust ring removal. *Canadian Journal of Ophthalmology* 2017;52(5):447-52.
261. Wong K, Stewart F. Competency-based training of basic trainees using human cadavers. *ANZ Journal of Surgery* 2004;74(8):639-42.
262. Rowland EB, Kleinert JM. Endoscopic carpal-tunnel release in cadavera. An investigation of the results of twelve surgeons with this training model. *JBJS* 1994;76(2):266-68.
263. Levine RL, Kives S, Cathey G, et al. The use of lightly embalmed (fresh tissue) cadavers for resident laparoscopic training. *Journal of Minimally Invasive Gynecology* 2006;13(5):451-6.
264. Bergeson RK, Schwend RM, DeLucia T, et al. How accurately do novice surgeons place thoracic pedicle screws with the free hand technique? *Spine* 2008;33(15):E501-7.
265. Tortolani PJ, Moatz BW, Parks BG, et al. Cadaver training module for teaching thoracic pedicle screw placement to residents. *Orthopedics* 2013;36(9):e1128-33.
266. Mowry SE, Hansen MR. Resident participation in cadaveric temporal bone dissection correlates with improved performance on a standardized skill assessment instrument. *Otology & Neurotology* 2013;35(1):77-83.

267. Awad Z, Tornari C, Ahmed S, et al. Construct validity of cadaveric temporal bones for training and assessment in mastoidectomy. *Laryngoscope* 2015;125(10):2376-81.
268. Egle JP, Malladi SVS, Gopinath N, et al. Simulation training improves resident performance in hand-sewn vascular and bowel anastomoses. *Journal of Surgical Education* 2015;72(2):291-96.
269. Nicandri GT, Cosgarea AJ, Hutchinson MR, et al. Assessing improvement in diagnostic knee arthroscopic skill during the AAOS fundamentals of knee and shoulder arthroscopy course for orthopaedic residents. *Orthopaedic Journal of Sports Medicine* 2015;3(7)
270. Kim SC, Fisher JG, Delman KA, et al. Cadaver-Based Simulation Increases Resident Confidence, Initial Exposure to Fundamental Techniques, and May Augment Operative Autonomy. *Journal of Surgical Education* 2016;73(6):e33-e41.
271. Martin KD, Patterson DP, Cameron KL. Arthroscopic Training Courses Improve Trainee Arthroscopy Skills: A Simulation-Based Prospective Trial. *Arthroscopy* 2016;32(11):2228-32.
272. Ciporen JN, Lucke-Wold B, Mendez G, et al. Endoscopic Management of Cavernous Carotid Surgical Complications: Evaluation of a Simulated Perfusion Model. *World Neurosurgery* 2017;98:388-96.
273. Martin M, Vashisht B, Frezza E, et al. Competency-based instruction in critical invasive skills improves both resident performance and patient safety. *Surgery* 1998;124(2):313-7.
274. Martin KD, Akoh CC, Amendola A, et al. Comparison of Three Virtual Reality Arthroscopic Simulators as Part of an Orthopedic Residency Educational Curriculum. *Iowa Orthop J* 2016;36:20-25.
275. Mednick Z, Tabanfar R, Alexander A, et al. Creation and validation of a simulator for corneal rust ring removal. *Canadian Journal of Ophthalmology* 2017;52(5):447-452.
276. Benkhadra M, Faust A, Ladoire S, et al. Comparison of fresh and Thiel's embalmed cadavers according to the suitability for ultrasound-guided regional anesthesia of the cervical region. *Surgical and Radiologic Anatomy* 2009;31(7):531.
277. Lloyd GM, Maxwell-Armstrong C, Acheson AG. Fresh frozen cadavers: an under-utilized resource in laparoscopic colorectal training in the United Kingdom. *Colorectal Disease* 2011;13(9):e303-e04.
278. Eisma R, Lamb C, Soames RW. From formalin to thiel embalming: What changes? One anatomy department's experiences. *Clinical Anatomy* 2013;26(5):564-71.
279. Gail M. Sullivan. Deconstructing Quality in Education Research. *Journal of Graduate Medical Education* 2011;3(2):121-24.
280. British Orthopaedic Trainees Association courses. <http://www.bota.org.uk/courses/> (accessed 03 Feb 2020).
281. Royal College of Surgeons of England: Courses to support every stage of your career. <https://www.rcseng.ac.uk/education-and-exams/courses/> (accessed 03 Feb 2020).
282. Robinson DA, Evans JH, Martellaro AJ, et al. Interactive translational research model and cadaveric simulation: where minimally invasive cardiac surgery and industry meet. *Annals of translational medicine* 2018;6(15):314-14.
283. Cecil J, Gupta A, Pirela-Cruz M. An advanced simulator for orthopedic surgical training. *International Journal of Computer Assisted Radiology and Surgery* 2018;13(2):305-19.

284. Logishetty K, Rudran B, Gofton W, et al. Total hip arthroplasty training in virtual reality transfers to the real world: a randomized blinded study. *Orthopaedic Proceedings* 2019;101-B(SUPP_6):27-27.
285. Specialist Advisory Committee. Trauma & Orthopaedic Curriculum Mapped to Simulation Options: British Orthopaedic Association, 2013.
286. Joint Committee on Surgical Training. JCST Quality Assurance Committee Meeting - 12th March 2018, 2018. <https://www.bota.org.uk/jcst-quality-assurance-committee-meeting-12th-march-2018/> (accessed 06 Feb 2020)
287. UK and Ireland In-Training Exam (UKITE): British Orthopaedic Association. <https://www.boa.ac.uk/learning-and-events/ukite.html> (accessed 06 Feb 2020).
288. Karam MD, Pedowitz RA, Natividad H, et al. Current and Future Use of Surgical Skills Training Laboratories in Orthopaedic Resident Education: A National Survey. *JBJS* 2013;95(1):e4.
289. Scott DJ. Patient Safety, Competency, and the Future of Surgical Simulation. *Simulation in Healthcare* 2006;1(3):164-70.
290. Tuijthof GJM, van Sterkenburg MN, Sierevelt IN, et al. First validation of the PASSPORT training environment for arthroscopic skills. *Knee Surgery, Sports Traumatology, Arthroscopy* 2010;18(2):218-24.
291. Angelo RL, Ryu RKN, Pedowitz RA, et al. Metric Development for an Arthroscopic Bankart Procedure: Assessment of Face and Content Validity. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2015;31(8):1430-40.
292. Colaco HB, Hughes K, Pearse E, et al. Construct Validity, Assessment of the Learning Curve, and Experience of Using a Low-Cost Arthroscopic Surgical Simulator. *Journal of Surgical Education* 2017;74(1):47-54.
293. Coughlin RP, Pauyo T, Sutton JC, 3rd, et al. A Validated Orthopaedic Surgical Simulation Model for Training and Evaluation of Basic Arthroscopic Skills. *The Journal of bone and joint surgery American volume* 2015;97(17):1465-71.
294. Cannon WD, Nicandri GT, Reinig K, et al. Evaluation of skill level between trainees and community orthopaedic surgeons using a virtual reality arthroscopic knee simulator. *The Journal of bone and joint surgery American volume* 2014;96(7):e57.
295. Rebolledo BJ, Hammann-Scala J, Leali A, et al. Arthroscopy skills development with a surgical simulator: a comparative study in orthopaedic surgery residents. *The American journal of sports medicine* 2015;43(6):1526-9.
296. Howells NR, Gill HS, Carr AJ, et al. Transferring simulated arthroscopic skills to the operating theatre. *The Journal of Bone and Joint Surgery British volume* 2008;90-B(4):494-99.
297. James HK. Measuring the educational impact of simulation training in Trauma & Orthopaedics. *Journal of Trauma & Orthopaedics* 2019;7(3):54-56.
298. Garfjeld Roberts P, Alvand A, Gallieri M, et al. Objectively Assessing Intraoperative Arthroscopic Skills Performance and the Transfer of Simulation Training in Knee Arthroscopy: A Randomized Controlled Trial. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2019;35(4):1197-209.e1.
299. Mitchell EL, Arora S, Moneta GL, et al. A systematic review of assessment of skill acquisition and operative competency in vascular surgical training. *Journal of Vascular Surgery* 2014;59(5):1440-55.
300. Okoro T, Sirianni C, Brigden D. The Concept of Surgical Assessment: Part 1 – Introduction. *The Bulletin of the Royal College of Surgeons of England* 2010;92(9):322-23.

301. Okoro T, Sirianni C, Brigden D. The Concept of Surgical Assessment: Part 2 – Available Tools. *The Bulletin of the Royal College of Surgeons of England* 2010;92(9):324-26.
302. Memon MA, Brigden D, Subramanya MS, et al. Assessing the surgeon's technical skills: analysis of the available tools. *Acad Med* 2010;85(5):869-80.
303. Van Der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education* 1996;1(1):41-67.
304. Bartlett JD, Lawrence JE, Stewart ME, et al. Does virtual reality simulation have a role in training trauma and orthopaedic surgeons? *The Bone & Joint Journal* 2018;100-B(5):559-65.
305. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349.
306. An VVG, Mirza Y, Mazomenos E, et al. Arthroscopic simulation using a knee model can be used to train speed and gaze strategies in knee arthroscopy. *The Knee* 2018;25(6):1214-21.
307. Angelo RL, Ryu RKN, Pedowitz RA, et al. A Proficiency-Based Progression Training Curriculum Coupled With a Model Simulator Results in the Acquisition of a Superior Arthroscopic Bankart Skill Set. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2015;31(10):1854-71.
308. Bhattacharyya R, Davidson DJ, Sugand K, et al. Knee Arthroscopy Simulation: A Randomized Controlled Trial Evaluating the Effectiveness of the Imperial Knee Arthroscopy Cognitive Task Analysis (IKACTA) Tool. *JBJS* 2017;99(19):e103.
309. Çetinkaya E, Çift H, Aybar A, et al. The timing and importance of motor skills course in knee arthroscopy training. *Acta Orthopaedica et Traumatologica Turcica* 2017;51(4):273-77.
310. Chong ACM, Pate RC, Prohaska DJ, et al. Validation of Improvement of Basic Competency in Arthroscopic Knot Tying Using a Bench Top Simulator in Orthopaedic Residency Education. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2016;32(7):1389-99. <https://doi.org/10.1016/j.arthro.2016.01.059>
311. Gomoll AH, Pappas G, Forsythe B, et al. Individual Skill Progression on a Virtual Reality Simulator for Shoulder Arthroscopy: A 3-Year Follow-up Study. *The American Journal of Sports Medicine* 2008;36(6):1139-42.
312. Jackson WFM, Khan T, Alvand A, et al. Learning and Retaining Simulated Arthroscopic Meniscal Repair Skills. *JBJS* 2012;94(17):e132.
313. Martin KD, Patterson D, Phisitkul P, et al. Ankle Arthroscopy Simulation Improves Basic Skills, Anatomic Recognition, and Proficiency During Diagnostic Examination of Residents in Training. *Foot & ankle international* 2015;36(7):827-35.
314. Rahm S, Wieser K, Bauer DE, et al. Efficacy of standardized training on a virtual reality simulator to advance knee and shoulder arthroscopic motor skills. *BMC Musculoskeletal Disorders* 2018;19(1):150.
315. Cannon WD, Garrett WE, Jr., Hunter RE, et al. Improving residency training in arthroscopic knee surgery with use of a virtual-reality simulator. A randomized blinded study. *The Journal of bone and joint surgery American volume* 2014;96(21):1798-806.
316. Dunn JC, Belmont PJ, Lanzi J, et al. Arthroscopic Shoulder Surgical Simulation Training Curriculum: Transfer Reliability and Maintenance of Skill Over Time. *Journal of Surgical Education* 2015;72(6):1118-23.

317. Waterman BR, Martin K, Cameron KL, et al. Simulation Training Improves Surgical Proficiency and Safety During Diagnostic Shoulder Arthroscopy Performed by Residents. *Orthopedics* 2016;39:e.479-e.85.
318. Butler BA, Lawton CD, Burgess J, et al. Simulation-Based Educational Module Improves Intern and Medical Student Performance of Closed Reduction and Percutaneous Pinning of Pediatric Supracondylar Humeral Fractures. *JBJS* 2017;99(23):e128.
319. LeBlanc J, Hutchison C, Hu Y, et al. A comparison of orthopaedic resident performance on surgical fixation of an ulnar fracture using virtual reality and synthetic models. *JBJS(Am)*2013;95(9):e60-5.
320. Nousiainen MT, Omoto DM, Zingg PO, et al. Training Femoral Neck Screw Insertion Skills to Surgical Trainees: Computer-Assisted Surgery Versus Conventional Fluoroscopic Technique. *Journal of Orthopaedic Trauma* 2013;27(2):87-92.
321. Ruder JA, Turvey B, Hsu JR, et al. Effectiveness of a Low-Cost Drilling Module in Orthopaedic Surgical Simulation. *Journal of Surgical Education* 2017;74(3):471-76.
322. Sonnadara RR, Van Vliet A, Safir O, et al. Orthopedic boot camp: Examining the effectiveness of an intensive surgical skills course. *Surgery* 2011;149(6):745-49.
323. Sonnadara RR, Garbedian S, Safir O, et al. Orthopaedic Boot Camp II: Examining the retention rates of an intensive surgical skills course. *Surgery* 2012;151(6):803-07.
324. Sonnadara RR, Garbedian S, Safir O, et al. Toronto orthopaedic boot camp III: examining the efficacy of student-regulated learning during an intensive, laboratory-based surgical skills course. *Surgery* 2013;154(1):29-33.
325. Tonetti J, Vadcard L, Girard P, et al. Assessment of a percutaneous iliosacral screw insertion simulator. *Orthopaedics & Traumatology: Surgery & Research* 2009;95(7):471-77.
326. Xiang L, Zhou Y, Wang H, et al. Significance of Preoperative Planning Simulator for Junior Surgeons' Training of Pedicle Screw Insertion. *Clinical Spine Surgery* 2015;28(1):E25-E29.
327. Brusalis C, M., Lawrence J, T, R., Ranade S, et al. Can a Novel, Low-Cost Simulation Model be Used to Teach Anterior Cruciate Ligament Graft Preparation? *Journal of Pediatric Orthopedics* 2017;37(4):e277-e81.
328. Dwyer T, Slade Shantz J, Chahal J, et al. Simulation of Anterior Cruciate Ligament Reconstruction in a Dry Model. *The American Journal of Sports Medicine* 2015;43(12):2997-3004.
329. Dwyer T, Schachar R, Leroux T, et al. Performance Assessment of Arthroscopic Rotator Cuff Repair and Labral Repair in a Dry Shoulder Simulator. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2017;33(7):1310-18.
330. Escoto A, Trejos AL, Naish MD, et al. Force sensing-based simulator for arthroscopic skills assessment in orthopaedic knee surgery. *Studies in Health Technology & Informatics* 2012;173:129-35.
331. Fucentese SF, Rahm S, Wieser K, et al. Evaluation of a virtual-reality-based simulator using passive haptic feedback for knee arthroscopy. *Knee Surgery, Sports Traumatology, Arthroscopy* 2015;23(4):1077-85.
332. Gomoll AH, O'Toole RV, Czarnecki J, et al. Surgical Experience Correlates with Performance on a Virtual Reality Simulator for Shoulder Arthroscopy. *The American Journal of Sports Medicine* 2007;35(6):883-88.

333. Howells NR, Brinsden MD, Gill RS, et al. Motion Analysis: A Validated Method for Showing Skill Levels in Arthroscopy. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2008;24(3):335-42.
334. Insel A, Carofino B, Leger R, et al. The Development of an Objective Model to Assess Arthroscopic Performance. *JBJS* 2009;91(9):2287-95.
335. Khanduja V, Lawrence JE, Audenaert E. Testing the Construct Validity of a Virtual Reality Hip Arthroscopy Simulator. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2017;33(3):566-71.
336. Martin KD, Belmont PJ, Schoenfeld AJ, et al. Arthroscopic basic task performance in shoulder simulator model correlates with similar task performance in cadavers. *JBJS(Am)* 2011;93(21):e1271-5.
337. Martin KD, Cameron K, Belmont PJ, et al. Shoulder arthroscopy simulator performance correlates with resident and shoulder arthroscopy experience. *JBJS(Am)* 2012;94(21):e160.
338. McCarthy A, Harley P, Smallwood R. Virtual arthroscopy training: do the "virtual skills" developed match the real skills required? *Stud Health Technol Inform* 1999;62:221-27.
339. McCarthy AD, Moody L, Waterworth AR, et al. Passive haptics in a knee arthroscopy simulator: is it valid for core skills training? *Clinical orthopaedics and related research* 2006;442:13-20.
340. Pedowitz RA, Esch J, Snyder S. Evaluation of a virtual reality simulator for arthroscopy skills development. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2002;18(6):1-6.
341. Pollard TCB, Khan T, Price AJ, et al. Simulated hip arthroscopy skills: learning curves with the lateral and supine patient positions: a randomized trial. *JBJS(Am)* 2012;94(10):e68.
342. Rahm S, Germann M, Hingsammer A, et al. Validation of a virtual reality-based simulator for shoulder arthroscopy. *Knee Surgery, Sports Traumatology, Arthroscopy* 2016;24(5):1730-37.
343. Rose K, Pedowitz R. Fundamental Arthroscopic Skill Differentiation With Virtual Reality Simulation. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2015;31(2):299-305.
344. Srivastava S, Youngblood PL, Rawn C, et al. Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. *Journal of Shoulder and Elbow Surgery* 2004;13(2):196-205.
345. Tashiro Y, Miura H, Nakanishi Y, et al. Evaluation of Skills in Arthroscopic Training Based on Trajectory and Force Data. *Clinical Orthopaedics and Related Research* 2009;467(2):546-52.
346. Toftte JN, Westerlind BO, Martin KD, et al. Knee, Shoulder, and Fundamentals of Arthroscopic Surgery Training: Validation of a Virtual Arthroscopy Simulator. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2017;33(3):641-46.e3.
347. Wong IH, Denkers M, Urquhart N, et al. Construct validity testing of the Arthroscopic Knot Trainer (ArK). *Knee Surgery, Sports Traumatology, Arthroscopy* 2015;23(3):906-11.
348. Akhtar K, Sugand K, Sperrin M, et al. Training safer orthopedic surgeons. Construct validation of a virtual-reality simulator for hip fracture surgery. *Acta orthopaedica* 2015;86(5):616-21.
349. Aoude A, Alhamzah H, Fortin M, et al. The use of computer-assisted surgery as an educational tool for the training of orthopedic surgery residents in pedicle screw

- placement: a pilot study and survey among orthopedic residents. *Can J Surg* 2016;59(6):391-98.
350. Blyth P, Stott NS, Anderson IA. Virtual reality assessment of technical skill using the Bonedoc DHS simulator. *Injury* 2008;39(10):1127-33.
351. Christian MW, Griffith C, Schoonover C, et al. Construct Validation of a Novel Hip Fracture Fixation Surgical Simulator. *The Journal of the American Academy of Orthopaedic Surgeons* 2018;26(19):689-97.
352. Froelich JM, Milbrandt JC, Novicoff WM, et al. Surgical simulators and hip fractures: a role in residency training? *Journal of surgical education* 2011;68(4):298-302.
353. Garfjeld Roberts P, Guyver P, Baldwin M, et al. Validation of the updated ArthroS simulator: face and construct validity of a passive haptic virtual reality simulator with novel performance metrics. *Knee Surgery, Sports Traumatology, Arthroscopy* 2017;25(2):616-25.
354. Giurin I, Bréaud J, Rampal V, et al. A simulation model of nail bed suture and nail fixation: description and preliminary evaluation. *Journal of Surgical Research* 2018;228:142-46.
355. Hohn EA, Brooks AG, Leasure J, et al. Development of a Surgical Skills Curriculum for the Training and Assessment of Manual Skills in Orthopedic Surgical Residents. *Journal of Surgical Education* 2015;72(1):47-52.
356. Leong JJH, Leff DR, Das A, et al. Validation of orthopaedic bench models for trauma surgery. *JBJS(Br)* 2008;90-B(7):958-65.
357. Lopez G, Wright R, Martin D, et al. A cost-effective junior resident training and assessment simulator for orthopaedic surgical skills via fundamentals of orthopaedic surgery: AAOS exhibit selection. *The Journal of bone and joint surgery American volume* 2015;97(8):659-66.
358. Mayne IP, Brydges R, Mokhtar J, et al. Development and Assessment of a Distal Radial Fracture Model as a Clinical Teaching Tool. *JBJS* 2016;98(5):410-16.
359. Qassemayar Q, Boulart L. A 4-Task Skills Examination for Residents for the Assessment of Technical Ability in Hand Trauma Surgery. *Journal of Surgical Education* 2015;72(2):179-83.
360. Rambani R, Ward J, Viant W. Desktop-Based Computer-Assisted Orthopedic Training System for Spinal Surgery. *Journal of Surgical Education* 2014;71(6):805-09.
361. Shi J, Hou Y, Lin Y, et al. Role of Visuohaptic Surgical Training Simulator in Resident Education of Orthopedic Surgery. *World neurosurgery* 2018;111:e98-e104.
362. Sugand K, Wescott RA, Carrington R, et al. Teaching basic trauma: validating FluoroSim, a digital fluoroscopic simulator for guide-wire insertion in hip surgery. *Acta Orthopaedica* 2018;89(4):380-85.
363. Yehyawi TM, Thomas TP, Ohrt GT, et al. A simulation trainer for complex articular fracture surgery. *JBJS(Am)* 2013;95(132):e92.
364. Alvand A, Khan T, Al-Ali S, et al. Simple visual parameters for objective assessment of arthroscopic skill. *JBJS(Am)* 2012;94(13):e97.
365. Alvand A, Logishetty K, Middleton R, et al. Validating a Global Rating Scale to Monitor Individual Resident Learning Curves During Arthroscopic Knee Meniscal Repair. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2013;29(5):906-12.
366. Bayona S, Akhtar K, Gupte C, et al. Assessing Performance in Shoulder Arthroscopy: The Imperial Global Arthroscopy Rating Scale (IGARS). *JBJS(Am)* 2014;96(13):e112.

367. Dwyer T, Slade Shantz J, Kulasegaram KM, et al. Use of an Objective Structured Assessment of Technical Skill After a Sports Medicine Rotation. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2016;32(12):2572-81.e3.
368. Elliott MJ, Caprise PA, Henning AE, et al. Diagnostic Knee Arthroscopy: A Pilot Study to Evaluate Surgical Skills. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2012;28(2):218-24.
369. Koehler RJ, Nicandri GT. Using the arthroscopic surgery skill evaluation tool as a pass-fail examination. *JBJS(Am)* 2013;95(23):e1871-6.
370. Koehler RJ, Amsdell S, Arendt EA, et al. The Arthroscopic Surgical Skill Evaluation Tool (ASSET). *The American journal of sports medicine* 2013;41(6):1229-37.
371. Middleton RM, Baldwin MJ, Akhtar K, et al. Which Global Rating Scale? A Comparison of the ASSET, BAKSSS, and IGARS for the Assessment of Simulated Arthroscopic Skills. *JBJS(Am)* 2016;98(1):75-81.
372. Nwachukwu B, Gaudiani M, Hammann-Scala J, et al. A Checklist Intervention to Assess Resident Diagnostic Knee and Shoulder Arthroscopic Efficiency. *Journal of Surgical Education* 2017;74(1):9-15.
373. Olson T, Koehler R, Butler A, et al. Is there a Valid and Reliable Assessment of Diagnostic Knee Arthroscopy Skill? *Clinical Orthopaedics and Related Research®* 2013;471(5):1670-76.
374. Pedowitz RA, Nicandri GT, Angelo RL, et al. Objective Assessment of Knot-Tying Proficiency With the Fundamentals of Arthroscopic Surgery Training Program Workstation and Knot Tester. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2015;31(10):1872-79.
375. Phillips L, Cheung JJH, Whelan DB, et al. Validation of a Dry Model for Assessing the Performance of Arthroscopic Hip Labral Repair. *The American journal of sports medicine* 2017;45(9):2125-30.
376. Price AJ, Erturan G, Akhtar K, et al. Evidence-based surgical training in orthopaedics. *The Bone & Joint Journal* 2015;97-B(10):1309-15.
377. Slade Shantz JA, Leiter JR, Collins JB, et al. Validation of a Global Assessment of Arthroscopic Skills in a Cadaveric Knee Model. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2013;29(1):106-12.
378. Gallagher AG, Ryu RKN, Pedowitz RA, et al. Inter-rater Reliability for Metrics Scored in a Binary Fashion—Performance Assessment for an Arthroscopic Bankart Repair. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2018;34(7):2191-98.
379. Hodgins JL, Veillette C, Biau D, et al. The Knee Arthroscopy Learning Curve: Quantitative Assessment of Surgical Skills. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2014;30(5):613-21.
380. Hoyle AC, Whelton C, Umaar R, et al. Validation of a global rating scale for shoulder arthroscopy: a pilot study. *Shoulder & Elbow* 2012;4(1):16-21.
381. Koehler RJ, Goldblatt JP, Maloney MD, et al. Assessing Diagnostic Arthroscopy Performance in the Operating Room Using the Arthroscopic Surgery Skill Evaluation Tool (ASSET). *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2015;31(12):2314-19.e2.
382. Talbot CL, Holt EM, Gooding BWT, et al. The Shoulder Objective Practical Assessment Tool: Evaluation of a New Tool Assessing Residents Learning in Diagnostic Shoulder Arthroscopy. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2015;31(8):1441-49.

383. Anderson DD, Long S, Thomas GW, et al. Objective Structured Assessments of Technical Skills (OSATS) Does Not Assess the Quality of the Surgical Result Effectively. *Clinical Orthopaedics and Related Research* 2016;474(4):874-81.
384. Backstein D, Agnidis Z, Regehr G, et al. The effectiveness of video feedback in the acquisition of orthopedic technical skills. *American journal of surgery* 2004;187(3):427-32.
385. Bernard JA, Dattilo JR, Srikumaran U, et al. Reliability and Validity of 3 Methods of Assessing Orthopedic Resident Skill in Shoulder Surgery. *Journal of Surgical Education* 2016;73(6):1020-25.
386. Burns GT, King BW, Holmes JR, et al. Evaluating Internal Fixation Skills Using Surgical Simulation. *JBJS(Am)* 2017;99(5):e21.
387. MacEwan MJ, Dudek NL, Wood TJ, et al. Continued Validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): Use in the Simulated Environment. *Teaching and learning in medicine* 2016;28(1):72-9.
388. Pedersen P, Palm H, Ringsted C, et al. Virtual-reality simulation to assess performance in hip fracture surgery. *Acta orthopaedica* 2014;85(4):403-7.
389. Putnam MD, Kinnucan E, Adams JE, et al. On Orthopedic Surgical Skill Prediction—The Limited Value of Traditional Testing. *Journal of Surgical Education* 2015;72(3):458-70.
390. Williams JF, Watson SL, Baker DK, et al. Psychomotor Testing for Orthopedic Residency Applicants: A Pilot Study. *Journal of Surgical Education* 2017;74(5):820-27.
391. Van Heest A, Putnam M, Agel J, et al. Assessment of technical skills of orthopaedic surgery residents performing open carpal tunnel release surgery. *JBJS(Am)* 2009;91(12):2811-7.
392. VanHeest A, Kuzel B, Agel J, et al. Objective structured assessment of technical skill in upper extremity surgery. *The Journal of hand surgery* 2012;37(2):332-4.
393. Beard J, D., Marriott J, Purdie H, et al. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technol Assess* 2011;15(1):1-162.
394. Davies RM, Hadfield-Law L, Turner PG. Development and Evaluation of a New Formative Assessment of Surgical Performance. *Journal of Surgical Education* 2018;75(5):1309-16.
395. Gofton WT, Dudek NL, Wood TJ, et al. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Academic medicine : journal of the Association of American Medical Colleges* 2012;87(10):1401-7.
396. Hawkes DH, Harrison WJ. Critiquing operative fracture fixation: the development of an assessment tool. *European Journal of Orthopaedic Surgery & Traumatology* 2017;27(8):1083-88.
397. Hoffer MM, Hsu SC. Hand function in selection of orthopedics residents. *Academic Medicine* 1990;65(10):661.
398. Marriott J, Purdie H, Crossley J, et al. Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. *BJS* 2011;98(3):450-57.
399. Martin KD, Patterson DP, Cameron KL. Arthroscopic Training Courses Improve Trainee Arthroscopy Skills: A Simulation-Based Prospective Trial. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* 2016;32(11):2228-32.

400. Angelo RL, Ryu RKN, Pedowitz RA, et al. The Bankart Performance Metrics Combined With a Cadaveric Shoulder Create a Precise and Accurate Assessment Tool for Measuring Surgeon Skill. *Arthroscopy* 2015;31(9):1655-70.
401. Howells NR, Auplish S, Hand GC, et al. Retention of arthroscopic shoulder skills learned with use of a simulator. Demonstration of a learning curve and loss of performance level after a time delay. *JBJS(Am)* 2009;91(5):1207-13.
402. Karlen AI, Solberg EJ, Quanbeck DS, et al. Orthopaedic Surgery Residency Rotations and Correlation With Orthopaedic In-Training Examination Performance. *Journal of Surgical Education* 2018;75(5):1325-28.
403. Swanson D, Marsh JL, Hurwitz S, et al. Utility of AAOS OITE Scores in Predicting ABOS Part I Outcomes: AAOS Exhibit Selection. *JBJS* 2013;95(12):e84.
404. Bradley CS, Moktar J, Maxwell A, et al. A Reliable and Valid Objective Structured Assessment of Technical Skill for the Application of a Pavlik Harness Based on International Expert Consensus. *Journal of pediatric orthopedics* 2016;36(7):768-72.
405. Moktar J, Popkin CA, Howard A, et al. Development of a cast application simulator and evaluation of objective measures of performance. *JBJS(Am)* 2014;96(9):e76.
406. Angelo RL, Pedowitz RA, Ryu RKN, et al. The Bankart Performance Metrics Combined With a Shoulder Model Simulator Create a Precise and Accurate Training Tool for Measuring Surgeon Skill. *Arthroscopy* 2015;31(9):1639-54.
407. Rambani R, Viant W, Ward J, et al. Computer-Assisted Orthopedic Training System for Fracture Fixation. *Journal of Surgical Education* 2013;70(3):304-08.
408. Blyth P, Stott NS, Anderson IA. A simulation-based training system for hip fracture fixation for use within the hospital environment. *Injury* 2007;38(10):1197-203.
409. Norcini J, Anderson B, Bollela V, et al. Criteria for good assesment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher* 2011;33:206-14.
410. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, et al. Objective assessment of technical surgical skills. *British Journal of Surgery* 2010;97:972-87.
411. Hammond I, Karthigasu K. Training, assessment and competency in gynaecologic surgery. *Best Practice & Research Clinical Obstetrics and Gynaecology* 2006;20:173-87.
412. AO Foundation. <https://www.aofoundation.org/Structure/Pages/default.aspx>. (accessed 06 Feb 2020).
413. Tsagakataki M, Choudhary A. Mersey deanery ophthalmology trainees' views of the objective assessment of surgical and technical skills (OSATS) workplace-based assessment tool. *Perspectives on Medical Education* 2013;2(1):21-27.
414. Baumgaertner MR, Curtin SL, Lindskog DM, et al. The value of the tip-apex distance in predicting failure of fixation of peritrochanteric fractures of the hip. *JBJS(Am)* 1995;77:1058-64.
415. Hunter AR, Baird E, Reed M. Procedure-based assessmenrs in trauma and orthopaedic training - The trainees' perspective. *Medical Teacher* 2015;37:444-49.
416. Pitts D, Rowley DI, Sher JL. Assessment of performance in orthopaedic training. *JBJS(Br)* 2005;87-B(9):1187-91.
417. Agha RA, Fowler AJ, Sevdalis N. The role of non-technical skills in surgery. *Ann Med Surg* 2015;4(4):422-27.

418. Costa ML, Achten J, Parsons NR, et al. Percutaneous fixation with Kirschner wires versus volar locking plate fixation in adults with dorsally displaced fracture of distal radius: randomised controlled trial. *BMJ* 2014;349:g4807.
419. Wilson I, Bohm E, Lübbecke A, et al. Orthopaedic registries with patient-reported outcome measures. *EFORT Open Reviews* 2019;4(6):357-67.
420. Baker PN, Deehan DJ, Lees D, et al. The effect of surgical factors on early patient-reported outcome measures (PROMS) following total knee replacement. *JBJS(Br)* 2012;94-B(8):1058-66.
421. Vincent HK, Hagen JE, Zdziarski-Horodyski LA, et al. Patient-Reported Outcomes Measurement Information System Outcome Measures and Mental Health in Orthopaedic Trauma Patients During Early Recovery. *Journal of Orthopaedic Trauma* 2018;32(9):467-73.
422. Hamilton DF, Lane JV, Gaston P, et al. What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. *BMJ Open* 2013;3(4):e002525.
423. Donohoe H, Stellefson M, Tennant B. Advantages and Limitations of the e-Delphi technique: Implications for Health Education Researchers. *American Journal of Health Education* 2012;13(1):38-46.
424. Black N, Murphy M, Lamping D, et al. Consensus Development Methods: A Review of Best Practice in Creating Clinical Guidelines. *J Health Serv Res Policy* 1999;236(4):236-48.
425. Singh P, Aggarwal R, Zevin B, et al. A Global Delphi Consensus Study on Defining and Measuring Quality in Surgical Training. *Journal of the American College of Surgeons* 2014;219(3):346-53.e7.
426. Schneider P, Evaniew N, Rendon JS, et al. Moving forward through consensus: protocol for a modified Delphi approach to determine the top research priorities in the field of orthopaedic oncology. *BMJ Open* 2016;6(5):e011780.
427. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology* 2014;67(4):401-09.
428. Keene DJ, Mistry D, Nam J, et al. The Ankle Injury Management (AIM) trial: a pragmatic, multicentre, equivalence randomised controlled trial and economic evaluation comparing close contact casting with open surgical reduction and internal fixation in the treatment of unstable ankle fractures in patients aged over 60 years. *Health Technol Assess* 2016;20(75).
429. Kearney RS, McKeown R, Stevens S, et al. Cast versus functional brace in the rehabilitation of patients treated for an ankle fracture: protocol for the UK study of ankle injury rehabilitation (AIR) multicentre randomised trial. *BMJ Open* 2018;8(12):e027242.
430. Vanrusselt J, Vansevenant M, Vanderschueren G, et al. Postoperative radiograph of the hip arthroplasty: what the radiologist should know. *Insights into Imaging* 2015;6(6):591-600.
431. McBride TJ, Prakash D. How to read a postoperative total hip replacement radiograph. *Postgraduate Medical Journal* 2011;87(1024):101.
432. Munuera L, Garcia-Cimbrelo E. The femoral component in low-friction arthroplasty after ten years. *Clinical orthopaedics and related research* 1992(279):163-75.
433. Barrack RL, Mulroy RD, Harris WH. Improved cementing techniques and femoral component loosening in young patients with hip arthroplasty. A 12-year radiographic review. *JBJS(Br)* 1992;74-B(3):385-89.

434. Mont MA, Sedlin ED, Weiner LS, et al. Postoperative radiographs as predictors of clinical outcome in unstable ankle fractures. *Journal of orthopaedic trauma* 1992;6(3):352-57.
435. Pettrone FA, Gail M, Pee D, et al. Quantitative criteria for prediction of the results after displaced fracture of the ankle. *JBJS(Am)* 1983;65(5):667-77.
436. Joy G, Patzakis MJ, Harvey JP. Precise Evaluation of the Reduction of Severe Ankle Fractures: Technique and Correlation with End Results. *JBJS(Am)* 1974;56(5):979-93.
437. Phillips WA, Schwartz HS, Keller CS, et al. A prospective, randomized study of the management of severe ankle fractures. *JBJS(Am)* 1985;67(1):67-78.
438. Salem KMI, Shannak OA, Scammell BE, et al. Predictors and outcomes of treatment in hip hemiarthroplasty dislocation. *Annals of the Royal College of Surgeons of England* 2014;96(6):446-51.
439. Kreimeier U. Pathophysiology of fluid imbalance. *Crit Care* 2000;4 Suppl 2(Suppl 2):S3-S7.
440. van Kraaij DJ, Jansen RW, Hoefnagels W. Monitoring hypovolemia in healthy elderly subjects by measuring blood pressure response to Valsalva's maneuver. *Geriatric Nephrology and Urology* 1999;9(2):73-79.
441. Goodnough LT. Risks of blood transfusion. *Critical Care Medicine* 2003;31(12):S678-S86.
442. Varney SJ, Guest JF. The annual cost of blood transfusions in the UK. *Transfusion Medicine* 2003;13(4):205-18.
443. Ghadimi K, Levy JH, Welsby IJ. Perioperative management of the bleeding patient. *BJA: British Journal of Anaesthesia* 2016;117(suppl_3):iii18-iii30.
444. Sambandam B, Batra S, Gupta R, et al. Blood conservation strategies in orthopedic surgeries: A review. *J Clin Orthop Trauma* 2013;4(4):164-70.
445. Gao F-Q, Li Z-J, Zhang K, et al. Four Methods for Calculating Blood-loss after Total Knee Arthroplasty. *Chinese Medical Journal* 2015;128(21):2856-60.
446. Seol Y-J, Seon J-K, Lee S-H, et al. Effect of Tranexamic Acid on Blood Loss and Blood Transfusion Reduction after Total Knee Arthroplasty. *Knee Surg Relat Res* 2016;28(3):188-93.
447. Kalairajah Y, Simpson D, Cossey AJ, et al. Blood loss after total knee replacement. *JBJS(Br)* 2005;87-B(11):1480-82.
448. Kanonidou Z, Karystianou G. Anesthesia for the elderly. *Hippokratia* 2007;11(4):175-77.
449. James HK, Chapman AW, Pattison GTR, et al. Current status of Cadaveric Simulation for surgical training: a Systematic Review. *British Journal of Surgery* 2019;106(13):1726-1734.
450. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
451. Training Standards Committee of the British Orthopaedic Association. Specialist Training in Trauma and Orthopaedics - 2015 Curriculum (2018 Update). London: General Medical Council, 2018:1-232. https://www.gmc-uk.org/-/media/documents/T_and_O_inc._Trauma_TIG_approved_Jul_17.pdf_72511367.pdf (accessed 06 Feb 2020)
452. eLogbook - the Pan-Surgical Electronic Logbook for the United Kingdom & Ireland. <https://www.elogbook.org/>. (accessed 06 Feb 2020)
453. The National Hip Fracture Database. <https://www.nhfd.co.uk/> (accessed 06 Feb 2020)

454. Mayhew D, Mendonca V, Murthy BVS. A review of ASA physical status – historical perspectives and modern developments. *Anaesthesia* 2019;74(3):373-79.
455. Science Care. <https://www.sciencecare.com/> (accessed 06 Feb 2020)
456. Kaplan DJ, Patel JN, Liporace FA, et al. Intraoperative radiation safety in orthopaedics: a review of the ALARA (As low as reasonably achievable) principle. *Patient Safety in Surgery* 2016;10(1):27.
457. Herscovici DJ, Sanders RW. The Effects, Risks, and Guidelines for Radiation Use in Orthopaedic Surgery. *Clinical Orthopaedics and Related Research* 2000;375:126-32.
458. Rubio-Avila J, Madden K, Simunovic N, et al. Tip to apex distance in femoral intertrochanteric fractures: a systematic review. *Journal of Orthopaedic Science* 2013;18(4):592-98.
459. Hsueh K-K, Fang C-K, Chen C-M, et al. Risk factors in cutout of sliding hip screw in intertrochanteric fractures: an evaluation of 937 patients. *Int Orthop* 2010;34(8):1273-76.
460. Schwartzmann CR, Jacobus LS, Spinelli LdF, et al. Dynamic hip screw for the treatment of femoral neck fractures: a prospective study with 96 patients. *ISRN Orthop* 2014;2014:257871-71.
461. Andruszkow H, Frink M, Frömke C, et al. Tip apex distance, hip screw placement, and neck shaft angle as potential risk factors for cut-out failure of hip screws after surgical treatment of intertrochanteric fractures. *Int Orthop* 2012;36(11):2347-54.
462. Konyves A, Bannister GC. The importance of leg length discrepancy after total hip arthroplasty. *JBJS(Am)* 2005;87-B(2):155-57.
463. Devitt A, O'Sullivan T, Quinlan W. 16- to 25-year follow-up study of cemented arthroplasty of the hip in patients aged 50 years or younger. *The Journal of Arthroplasty* 1997;12(5):479-89.
464. Kobayashi S, Eftekhari NS, Terayama K. Predisposing factors in fixation failure of femoral prostheses following primary Charnley low friction arthroplasty. A 10- to 20-year followup study. *Clinical orthopaedics and related research* 1994(306):73-83.
465. Enke O, Levy YD, Bruce WJM. Accuracy of leg length and femoral offset restoration after total hip arthroplasty with the utilisation of an intraoperative calibration gauge. *HIP International* 2019;1120700019836383.
466. Lecerf G, Fessy MH, Philippot R, et al. Femoral offset: Anatomical concept, definition, assessment, implications for preoperative templating and hip arthroplasty. *Orthopaedics & Traumatology: Surgery & Research* 2009;95(3):210-19.
467. Chambers IR, Fender D, McCaskie AW, et al. Radiological features predictive of aseptic loosening in cemented Charnley femoral stems. *JBJS(Br)* 2001;83-B(6):838-42.
468. Said GZ, Farouk O, El-Sayed A, et al. Salvage of failed dynamic hip screw fixation of intertrochanteric fractures. *Injury* 2006;37(2):194-202.
469. Hammad A, Abdel-Aal A, Said HG, et al. Total hip arthroplasty following failure of dynamic hip screw fixation of fractures of the proximal femur. *Acta Orthop Belg* 2008;74(6):788-92.
470. Dittmer DK, Teasell R. Complications of immobilization and bed rest. Part 1: Musculoskeletal and cardiovascular complications. *Can Fam Physician* 1993;39:1428-37.
471. Kunutsor SK, Barrett MC, Beswick AD, et al. Risk factors for dislocation after primary total hip replacement: a systematic review and meta-analysis of 125 studies

- involving approximately five million hip replacements. *The Lancet Rheumatology* 2019;1(2):e111-e21.
472. Jerosch-Herold C. An evidence-based approach to choosing outcome measures: a checklist for the critical appraisal of validity, reliability and responsiveness studies. *British Journal of Occupational Therapy* 2005;68(5):347-53.
473. National Institute for Health and Care Excellence. Hip fracture: management. Clinical guideline (CG124), 2017. <https://www.nice.org.uk/Guidance/CG124> (accessed 06 Feb 2020)
474. Cleland J, Walker KG, Gale M, et al. Simulation-based education: understanding the socio-cultural complexity of a surgical training ‘boot camp’. *Medical Education* 2016;50(8):829-41.
475. Braun V, Clarke C. Ten fundamentals of qualitative research. Successful Qualitative Research - A practical guide for beginners. 1 ed: SAGE Publications 2013:19-40.
476. Ng S, Lingard L, Kennedy T. Qualitative research in medical education: Methodologies and methods. In: Swanwick T, ed. Understanding Medical Education: Evidence, Theory and Practice. 2 ed: Wiley-Blackwell 2013:371-84.
477. Braun V, Clark C. Successfully collecting qualitative data. Interactive data collection 1: interviews. Successful Qualitative Research - A practical guide for beginners. 1 ed: SAGE Publications 2013:77-105.
478. Braun V, Clark C. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006;3(2):77-101.
479. Braun V, Clark C. Successfully analysing qualitative data. First analytic steps: familiarisation and data coding. Successful Qualitative Research - A practical guide for beginners. 1 ed: SAGE Publications 2013:201-22.
480. NVivo qualitative data analysis Software; QSR International Pty Ltd. Version 11, 2017.
481. Braun V, Clark C. Identifying patterns across data. Successful Qualitative Research - A practical guide for beginners: SAGE Publications 2013:223-46.
482. Braun V, Clark C. Analysing and interpreting patterns across data. Successful Qualitative Research - A practical guide for beginners: SAGE Publications 2013:248-73.
483. Haroon M, Yasin F, Eckel R, et al. Perceptions and attitudes of hospital staff toward paging system and the use of mobile phones. *International Journal of Technology Assessment in Health Care* 2010;26(4):377-81.
484. Blakey JD, Guy D, Simpson C, et al. Multimodal observational assessment of quality and productivity benefits from the implementation of wireless technology for out of hours working. *BMJ Open* 2012;2(2):e000701.
485. Roueche A, Hewitt J. ‘Wading through treacle’: quality improvement lessons from the frontline. *BMJ Quality & Safety* 2012;21(3):179.
486. Prentice R. Drilling Surgeons: The Social Lessons of Embodied Surgical Learning. *Science, Technology, & Human Values* 2007;32(5):534-53.
487. Jensen RD, Seyer-Hansen M, Cristancho SM, et al. Being a surgeon or doing surgery? A qualitative study of learning in the operating room. *Medical Education* 2018;52(8):861-76.
488. Braun V, Clark C. Quality criteria and techniques for qualitative research. Successful Qualitative Research - A practical guide for beginners: SAGE Publications 2013:277-94.
489. Schoonenboom J, Johnson RB. How to Construct a Mixed Methods Research Design. *Kolner Z Soz Sozpsychol* 2017;69(Suppl 2):107-31.

490. Kheir E, Charopoulos I, Dimitriou R, et al. The Health Economics of Ankle Fracture Fixation. *The Bulletin of the Royal College of Surgeons of England* 2012;94(4):1-5.
491. All types of airplanes from jumbo jets to small planes. <https://www.aircraftcompare.com/blog/types-of-airplanes/> (accessed 06 Feb 2020)
492. Logishetty K, Rudran B, Cobb JP. Virtual reality training improves trainee performance in total hip arthroplasty: a randomized controlled trial. *The Bone & Joint Journal* 2019;101-B(12):1585-92.
493. Sugand K, Mawkin M, Gupte C. Training effect of using Touch Surgery™ for intramedullary femoral nailing. *Injury* 2016;47(2):448-52.
494. Haubruck P, Nickel F, Ober J, et al. Evaluation of App-Based Serious Gaming as a Training Method in Teaching Chest Tube Insertion to Medical Students: Randomized Controlled Trial. *J Med Internet Res* 2018;20(5):e195-e95.
495. Mulligan EP. Evaluation and management of ankle syndesmosis injuries. *Physical Therapy in Sport* 2011;12(2):57-69.
496. Hsu C-E, Huang K-C, Lin T-C, et al. Integrated risk scoring model for predicting dynamic hip screw treatment outcome of intertrochanteric fracture. *Injury* 2016;47(11):2501-06.
497. Synthes®. DHS/DCS System Surgical Technique. Switzerland, 2017. <https://orto.hi.is/skrar/27.pdf> (accessed 06 Feb 2020)
498. Reich SM, Jaffe WL, Kummer FJ. Biomechanical determination of the optimal number of fixation screws for the sliding hip screw plate. *Bull Hosp Jt Dis* 1993;53(4):43-44.

11. Appendix

11.1 Chapter 3: Data extraction form

DATA EXTRACTION FORM - SYSTEMATIC REVIEW OF CADAVERIC SIMULATION IN POSTGRADUATE SURGICAL TRAINING

Initial Eligibility Screening

FIRST AUTHOR	COUNTRY	YEAR

CRITERION	YES	NO
P = Doctors in training		
I = Cadaveric Simulation Training exposure		
O =An attempt at assessment of training exposure		
Full text available in English Language		

If any of the above are 'NO' - don't proceed with review, record as 'reject', with reason;

Methodological Rigour (MERSQI Score) please circle relevant;

Domain	Item	Item Score
Study Design	1.Study Design	
	Single group cross-sectional or single group post-test only	1
	Single group pre and post-test	1.5
	Non-randomized, 2 group	2
	Randomized controlled experiment	3
Sampling	2. Institutions	

	Single institution	0.5
	Two institutions	1
	More than 2 institutions	1.5
	3. Response Rate	
	Not applicable	n/a
	Response rate <50% or not reported	0.5
	Response rate 50-74%	1
	Response rate \geq 75%	1.5
<hr/>		
Type of Data	4. Type of Data	
	Assessment by study subject	0
	Objective measurement	1
<hr/>		
Validity of Evaluation Instruments' Scores	Not applicable	n/a
	5. Internal Structure	
	Not reported	0
	Reported	1
	6. Content	
	Not reported	0
	Reported	1
	7. Relationships to other variables	
	Not reported	0
	Reported	1
<hr/>		
Data Analysis	8. Appropriateness of analysis	
	Data analysis inappropriate for study design or type of data	0
	Data analysis appropriate for study design and type of data	1
	9. Sophistication of analysis	
	Descriptive analysis only	1
	Beyond descriptive analysis	2
<hr/>		
Outcome	10. Outcome	
	Satisfaction, attitudes, perceptions, opinions, general facts	1
	Knowledge, skills	1.5
	Behaviours	2
	Patient/health care outcome	3
<hr/>		
Total		
<hr/>		

11.2 Chapter 4: Dashboard survey questions

1. Is simulation training formally timetabled within the teaching curriculum? (yes/no)

- a. If **yes** – details of type of simulation offered by stage of training (please tick all that apply)

Simulation Modality	Early CT1-2	Mid ST3- ST5	Advanced ST6-ST8	Not offered
Box-trainer, low-realism (e.g. sawbones)				
Virtual reality (e.g. arthroscopy)				
Animal Tissue				
Human Cadaveric – anatomy/approaches only				
Human Cadaveric – whole procedures including implants and instruments				
Other (please specify)				

- b. If **no** – why not? (please select all that apply)
- i. Funding issues
 - ii. Lack of facilities
 - iii. Logistical difficulties with timetabling simulation teaching
 - iv. Lack of available trainers
 - v. Lack of trainee enthusiasm for simulation training
 - vi. Other (please specify – free text answer)

2. source of funding for simulation training (select all that apply);

- a. centralized (deanery/HEE)
- b. top-sliced/compulsory allocation of trainee study budget
- c. external source
- d. industry
- e. trainee self- funded
- f. Other (please specify – free text answer)

3. How do you measure the effect of simulation training? (select all that apply)

- a. Questionnaire feedback
- b. Participant knowledge test
- c. ISCP workplace based assessments (PBA, DOPS, CEX etc.)
- d. Other workplace based assessments (e.g. GOSLE)
- e. Objective performance outcome measures (e.g. operative time, error rate, hand-motion analysis etc.)
- f. Patient outcome measures

g. Other (please specify – free text answer)

4. **Would you be willing to be contacted by a research fellow if further details on simulation activity are required (yes/no)**

11.3 Chapter 5: Data extraction form

Assessment of technical skills and operative competence in orthopaedics – systematic review

Eligibility screening

FIRST AUTHOR	COUNTRY	YEAR

Inclusion criteria

CRITERION	YES	NO
Primary empirical research		
Assessment of open or arthroscopic orthopaedic surgical skills		
Assessment in simulated and/or live theatre		
Full text available in English Language		

Any 'NO' record as reject with reason.....

Title	
Study Aim	
Setting	
Assessment format	
Participants (number and stage of training)	
Skills Assessed	

Assessment tool or metrics	
Category*	
Succinct summary of study results	
Take home message related to assessment tool	

* Where evidence is present in the study, add to the sections ‘characteristics’, ‘strengths’, ‘limitations’ and utility index domains in the excel spreadsheet as appropriate.

11.4 Chapter 6: Consensus setting survey questions

Round 1



Copy of Can we measure the technical success of a trauma operation from the post-op x-r

Hemiarthroplasty

Please indicate if you consider the following measures to be important for assessing the technical success of a hemiarthroplasty operation. Please assume you are looking at an adequate standard AP post-op x-ray.

1. Leg length discrepancy

- Yes - by measuring alignment of the centre of the femoral head with the greater trochanter
- Yes - by measuring the distance between the acetabular teardrop and the lesser trochanter
- No

Other (please specify)

2. Femoral stem alignment (varus/valgus/neutral)

- Yes
- No

3. Cement mantle thickness

- Yes
- No

4. Where would you measure the cement mantle thickness?

- All 7 Gruen zones
- Metaphyseal zones (1 and 7)
- I do not measure

Other (please specify)

5. Is there anything else you would look for on a hemiarthroplasty x-ray to make a judgement on the technical success of the operation?



Copy of Can we measure the technical success of a trauma operation from the post-op x-r

Dynamic Hip Screw

Please indicate if you consider the following measures to be important for assessing the technical success of a DHS. Please assume you are looking at adequate, standard post-op AP and lateral X-rays.

6. Tip-apex distance

- Yes
 No

What do you consider an appropriate cut-off value for an acceptable TAD to be?

7. Fracture reduction pattern (trabecular angle)

- Yes
 No

Comments

8. Lag screw position in the head (Cleveland's 9 zones)

- Yes
 No

9. Plate flush with femur

- Yes
 No

10. Screws perpendicular with plate

- Yes
 No

11. Screws are all bicortical

Yes

No

12. Is there anything else you would look for on an x-ray of a DHS to make a judgement on the technical success of an operation?



Copy of Can we measure the technical success of a trauma operation from the post-op x-rays

Ankle ORIF

Please indicate if you consider the following measures to be important for assessing the technical success of an ankle ORIF. Please assume you are looking at adequate standard AP + lateral post-op x-rays

13. Talar tilt angle

Yes

No

14. Medial clear space

Yes

No

15. Medial malleolar displacement

Yes

No

16. Lateral malleolar displacement

Yes

No

17. Tibiofibular clear space

Yes

No

18. Talar subluxation

Yes

No

19. Is there anything else that you would look for on an x-ray of an ankle ORIF to make a judgement on the technical success of an operation?

Round 2



Can we measure the technical success of a trauma operation from the post-op x-ray - Round 2

Hemiarthroplasty

Getting consensus on the appropriate measures

1. 90% of respondents in round 1 agreed leg length discrepancy was relevant. Do you think a cut-off for acceptable LLD post-op of <10mm is reasonable?

- Yes
 No

Other (please specify)

2. 85% agreed that femoral stem alignment was important. The THA literature defines the ideal implant position as neutral compared to the femoral shaft axis, with varus/valgus malalignment defined as >3 degrees deviation from this. Is it reasonable to use this to define a technically successful hemi?

- Yes
 No

Other (please specify)

3. There was a 60% agreement split in favour of measuring cement mantle thickness. We propose a global visual assessment of mantle quality (i.e. are there any major lucencies or gaps) could be more useful than measuring thickness. Do you agree?

- Yes
 No

Other (please specify)

4. Is offset important in assessing the technical success?

Yes

No

Other (please specify)



Can we measure the technical success of a trauma operation from the post-op x-ray - Rou
2

DHS

Getting consensus on appropriate measures

5. 100% agreed tip apex distance is important. Do you think an acceptable cutoff for TAD is <25mm (controlled for magnification)

Yes

No

Other (please specify)

6. 60% of respondents agreed fracture reduction pattern was important. The literature says an AP reduction angle of 165-170 degrees is probably optimal, and <160 degrees (suggesting varus) can be considered a poor technical outcome in terms of predicting device failure. Do you think these are reasonable parameters?

Yes

No - I do not think reduction pattern is sufficiently important to warrant measuring

No - Reduction pattern is important but I would use other measures

Other (please specify)

7. Opinion was split on whether lag screw position in the head matters (55% in favour). We think it is probably worth measuring this as in the literature a high anterior screw position in the head seems to predict cut-out independent of the tip-apex distance. Do you think this is reasonable?

Yes

No

Other (please specify)



Can we measure the technical success of a trauma operation from the post-op x-ray - Round 2

Ankle ORIF

Getting consensus on the appropriate measures

8. 90% said medial clear space was important. We propose a cut-off of less than or equal to 4mm as the definition of acceptable. Is this reasonable?

- Yes
- No

Other (please specify)

9. 95% said medial and lateral malleolar displacement was important. We propose a cut-off of less than or equal to 2mm for both as acceptable limits. Is this reasonable?

- Yes
- No

Other (please specify)

10. 75% said tibiofibular clear space was important. We propose a cut-off of less than 5mm as the acceptable limit. Is this reasonable?

- Yes
- No

Other (please specify)

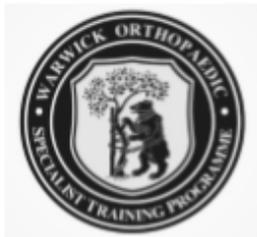
11. There was only 55% in favour of measuring talar tilt angle. We think maybe we should be measuring Talocrural angle instead as a more useful estimate of restoration of fibular length. Do you agree?

Yes

No

Other (please specify)

Round 3



Can we measure the technical success of a trauma operation from the post-op x-ray - Round 3

Hemiarthroplasty

Getting consensus on the appropriate measures

1. **Leg length discrepancy:** 67% of participants in round 2 thought 10mm was a reasonable cut-off for acceptable leg length discrepancy post-hemiarthroplasty. We need 75% majority agreement for consensus. We therefore propose revising the definition of acceptable LLD post-op to to <15mm. Is this reasonable?

Yes

No

Other (please specify)

2. **Stem alignment:** We need to define acceptable stem alignment. There was a 50:50 split on whether 3 degrees from neutral could reasonably be considered varus/valgus malalignment (we appreciate this has little/no practical correlation with clinical outcome, we are interested here in measuring the technical skill of the surgeon). We propose revising the cut-off to 5 degrees to be more generous. Is this reasonable?

Yes

No

Other (please specify)



Can we measure the technical success of a trauma operation from the post-op x-ray - Round 3

DHS

Getting consensus on appropriate measures

3. **Fracture reduction:** There was a 60:40 split in favour of measuring AP trabecular angle (the angle formed by the axis of femoral shaft and medial compressive trabeculae of femoral head). The anatomic range is 150-170 degrees, and research evidence suggests that varus malreduction significantly increases the likelihood of device failure ($p=0.002$, Hsu et al, 2016). We therefore propose a cut-off of 170 degrees as acceptable. Is this reasonable?

Yes

No

Other (please specify)



Can we measure the technical success of a trauma operation from the post-op x-ray - Round 3

Ankle ORIF

Getting consensus on the appropriate measures

4. **Malleolar displacement:** 71% agreed 2mm was an acceptable cut-off for medial and lateral malleolar displacement. The evidence supports this view, as a post-op medial/lateral malleolar displacement >2mm correlates with poor clinical outcome (Joy et al, 1974). We need 75% panel agreement to have achieved consensus. Do you agree with a 2mm cut-off?

- Yes
 No

Other (please specify)

5. **Tibiofibular clear space:** 67% of participants agreed tibiofibular clear space should be less than 5mm to be acceptable. (distance between the lateral border of the posterior tibial malleolus and the medial border of the fibula measured parallel to the tibial plafond, 1cm above it). This view is supported by the evidence as >5mm tibiofibular clear space after surgery predicts poor clinical outcome (Pettrone et al, 1983). Is this reasonable?

- Yes
 No

Other (please specify)

6. **Talo-crural angle:** 67% of participants think measuring talocrural angle is a useful measure of fibular length restoration. Normal is 80 degrees, and the evidence suggests that a post-op talocrural angle less than or equal to 5 degrees from normal predicts a good clinical outcome (Mont et al, 1992; Phillips et al, 1985). We accept as a limitation of the study that the quality of AP/ mortise views of the ankle in the trauma setting are variable, and need to work with what we have got. Do you agree that 5 degrees from normal is a reasonable cut-off point?

Yes

No

Other (please specify)

11.5 Chapter 7: CAD:TRAUMA radiology manual



cad:trauma study
INNOVATIVE SIMULATION TRAINING FOR SURGEONS

Radiology Manual



Ankle ORIF (adapted from AIR trial⁴²⁹)

Summary of cad:trauma radiology measurements

1. **Medial clear space**

- Definition: Distance between the lateral border of the medial malleolus and the medial border of the talus
- Parameter: $\leq 4\text{mm}$

2. **Medial malleolar displacement**

- Definition: The greatest step displacement on the articular surface
- Parameter: $\leq 2\text{mm}$

3. **Lateral malleolar displacement**

- Definition: The shortest distance across the greatest fracture gap anywhere between the two major fracture fragments
- Parameter: $\leq 2\text{mm}$

4. **Tibiofibular clear space**

- Definition: Distance between the lateral border of posterior tibial malleolus and medial border of the fibula measured parallel to the tibial plafond 1cm above it
- Parameter: $< 5\text{mm}$

5. **Talocrural angle**

- Definition: The angle between a line perpendicular to the plafond line and the line intersecting the tips of lateral and medial malleoli
- Parameter: $80^\circ \pm 5$ degrees

6. **Gross metalwork assessment**

- Definition: Global assessment of metalwork position
- Parameter: No intra-articular screws, plate flush with fibular

Explanation and supporting evidence (adapted from the AIR Trial⁴²⁹ radiology manual and modified per the cad:trauma Delphi exercise)

1. Medial clear space (Mont et al, 1992)

Patients with a medial clear space $\leq 4\text{mm}$ on the post-operative radiograph were significantly more likely to have a good/excellent functional outcome⁴³⁴

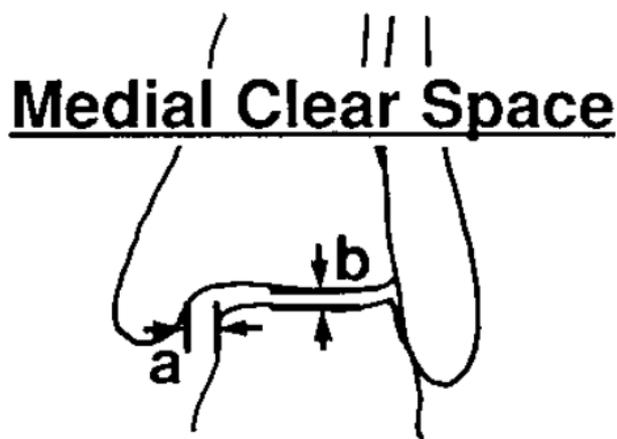


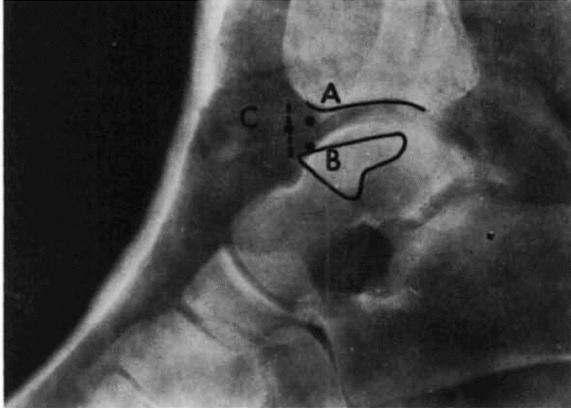
Figure adapted from Mont et al.

a = medial clear space. Should be $\leq 4\text{mm}$ and equal to the distance between the superior border of the talus and the articular surface of the tibia (b).

Care should be taken when doing this measurement to measure the distance between either the anterior borders of the medial malleolus and talus, or their respective posterior borders, and not to use to anterior border of one and the posterior border of the other⁴²⁹.

2. Medial malleolar displacement (Joy et al, 1974)

Medial malleolar displacement $>2\text{mm}$ post-reduction/surgery correlates with poor clinical outcome⁴³⁵



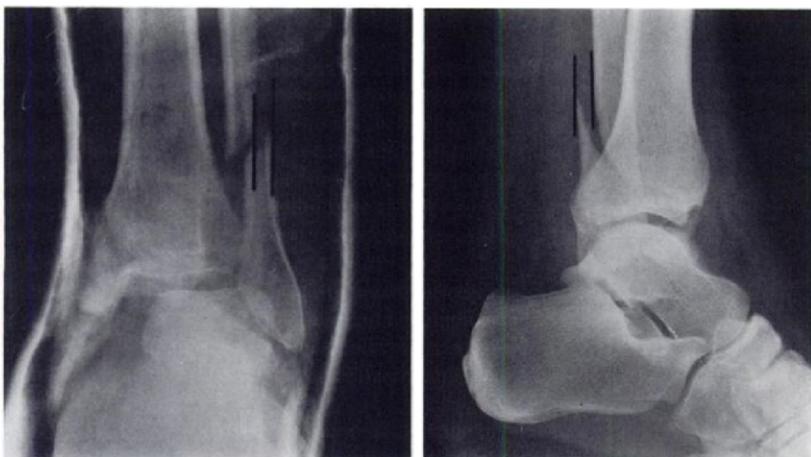
From Joy et al

The medial malleolar fragment is displaced distally (distance between A and B) and anteriorly (distance between A-C)

Distal and anterior displacement is best determined on the lateral radiograph
Lateral displacement is best determined on the AP view.

3. Lateral malleolar displacement (Joy et al, 1974)

Lateral malleolar displacement $>2\text{mm}$ post reduction/surgery correlates with poor clinical outcome⁴³⁵



From Joy et al

Antero-posterior displacement is measured on the lateral radiograph, medial-lateral displacement is measured on the AP view at the level of the fracture

5. Talo-crural angle (Mont et al, 1992; Phillips et al, 1985)

Talocrural angle $\leq 5^\circ$ from normal predicts good clinical outcome⁴³⁷

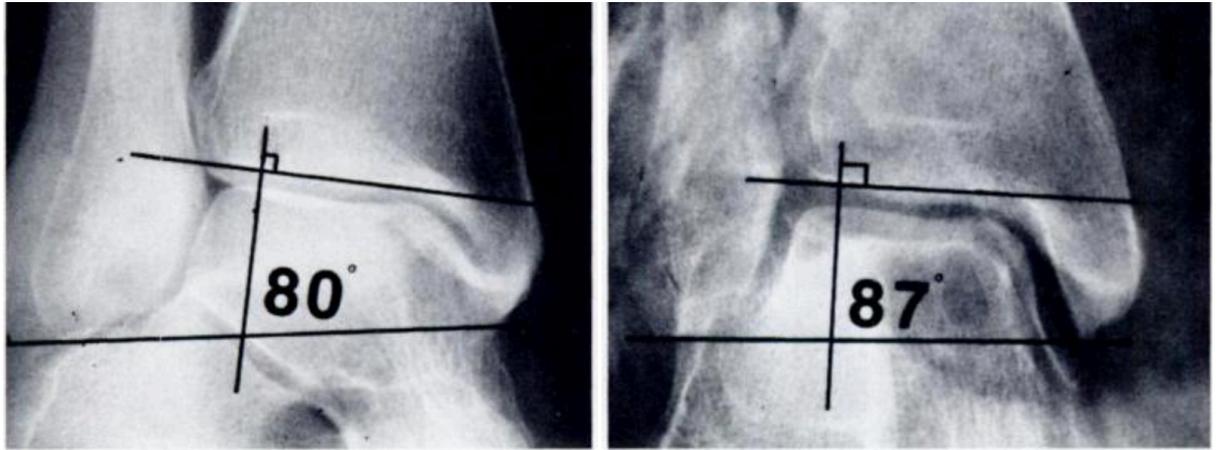


Figure from Phillips et al, 1985⁴³⁷

Talocrural angle is the superior and medial angle formed by the intersection of a line joining the tips of both malleoli and of a line perpendicular to the distal tibia articular surface⁴³⁷

Normal talo-crural angle suggests restoration of fibular length⁴³⁴

Normal = 80°

Dynamic Hip Screw (figures adapted from AO⁴¹²)

Summary of cad:trauma radiology measurements:

1. Tip-apex distance:

- Definition: Sum of the distance from the tip of the screw to the apex of the femoral head
- Parameter: <25mm

2. Trabecular angle:

- Definition: Angle formed by axis of femoral shaft and medial compressive trabeculae of femoral head
- Parameter:
 - i. Valgus reduction >170°
 - ii. Varus reduction <150°
 - iii. Anatomic reduction 150-170°

3. Lag screw position in femoral head:

- Definition: Femoral head divided into equal thirds on anteroposterior and lateral radiographs.
- Parameter:
 - i. AP view; Superior/Central/Inferior
 - ii. Lateral view; Posterior/Central/Anterior

4. Plate position:

- Definition: Position of plate against lateral femoral cortex
- Parameter:
 - i. AP view; Plate flush with cortex, no gaps seen

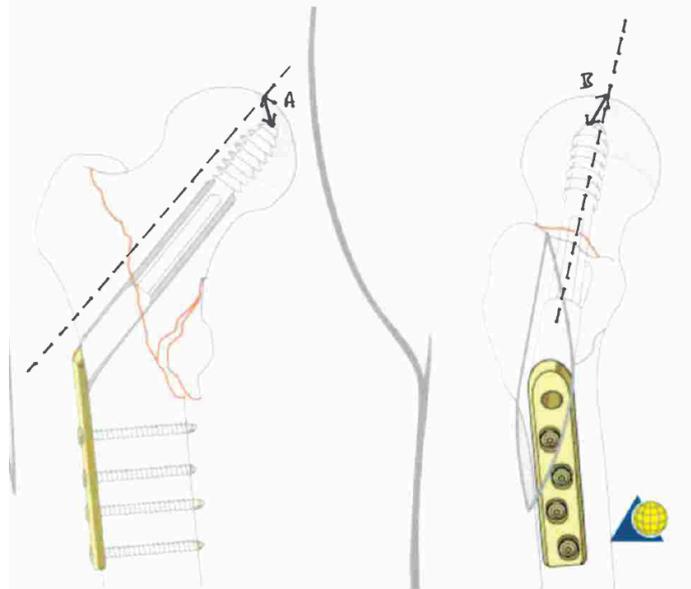
5. Cortical screw positions;

- Definition: Position of the cortical screws within femoral shaft
 - i. AP view; 8 cortex hold

Explanation and supporting evidence:

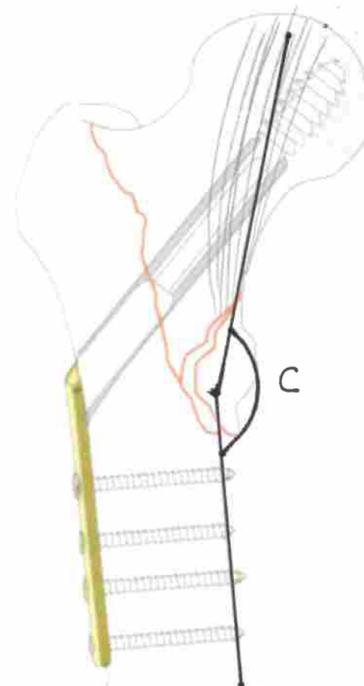
1. Tip-apex distance (Baumgaertner et al, 1995)

Tip-apex distance <25mm is significantly protective against cut-out and device failure⁴¹⁴



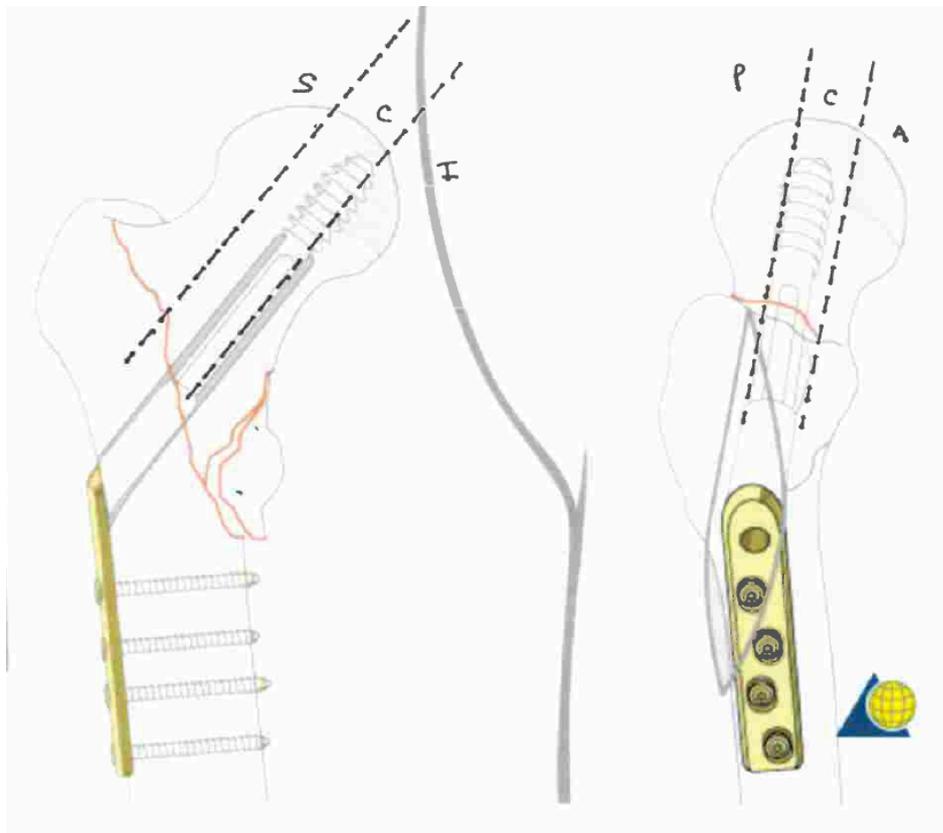
2. Trabecular angle (Hsu et al, 2016)

Varus mal-reduction significantly increases the likelihood of device failure (p=0.002)⁴⁹⁶



3. Lag screw position in femoral head (Pervez et al, 2016)

There is a significantly increased risk of cut-out for superior ($p=0.037$) or anteriorly ($p=0.0095$) placed screws



4. Plate position

A correctly positioned plate is flush with the lateral cortex with no gaps⁴⁹⁷.

5. Cortical screw positions (Reich et al, 1993)

Biomechanical laboratory testing of the tensile force on the plate fixation screws in a DHS showed that most of the force is borne by the proximal 3 screws. Four bicortical screws appeared adequate for fixation⁴⁹⁸

Hemiarthroplasty

Summary of cad:trauma radiology measurements:

1. Leg length discrepancy

- Definition: Difference in distance between femoral and pelvic reference lines on each side
- Parameter: $\leq 10\text{mm}$

2. Femoral stem alignment

- Definition: alignment of the stem with the longitudinal axis of the femoral shaft
- Parameter: degrees off neutral, varus/valgus

3. Cement mantle grade

- Definition: Completeness and thickness of the cement mantle that surrounds the femoral stem
- Parameter: Barrack grading A-D, thickness in mm

4. Femoral offset

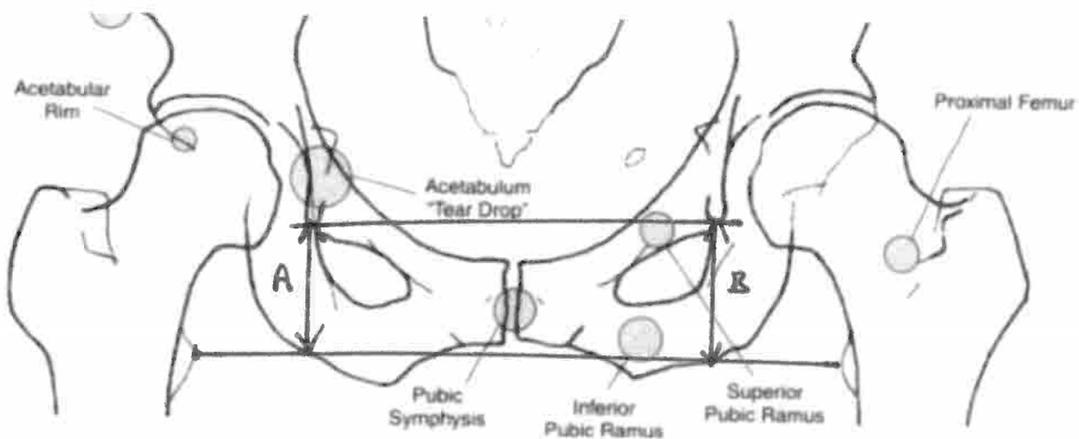
- Definition: The distance from the centre of rotation of the femoral head to a line bisecting the long axis of the femur
- Parameter: Usual range 41-44mm, should be equal to native side

Explanation and supporting evidence:

1. Leg length discrepancy (Vanrusselt et al, 2015)

Leg length discrepancy is common after hip (hemi)arthroplasty; a discrepancy of up to 1cm is well tolerated⁴³⁰

The leg length is measured by drawing a line transversely connecting the inferior borders of the acetabular tear drop, the pelvic reference line. The lesser trochanters are used as the femoral reference lines. Perpendicular lines are drawn from the pelvic reference line to the femoral reference lines, and the LLD is the difference between the two distances⁴³⁰.



LLD = Difference in distance between A and B

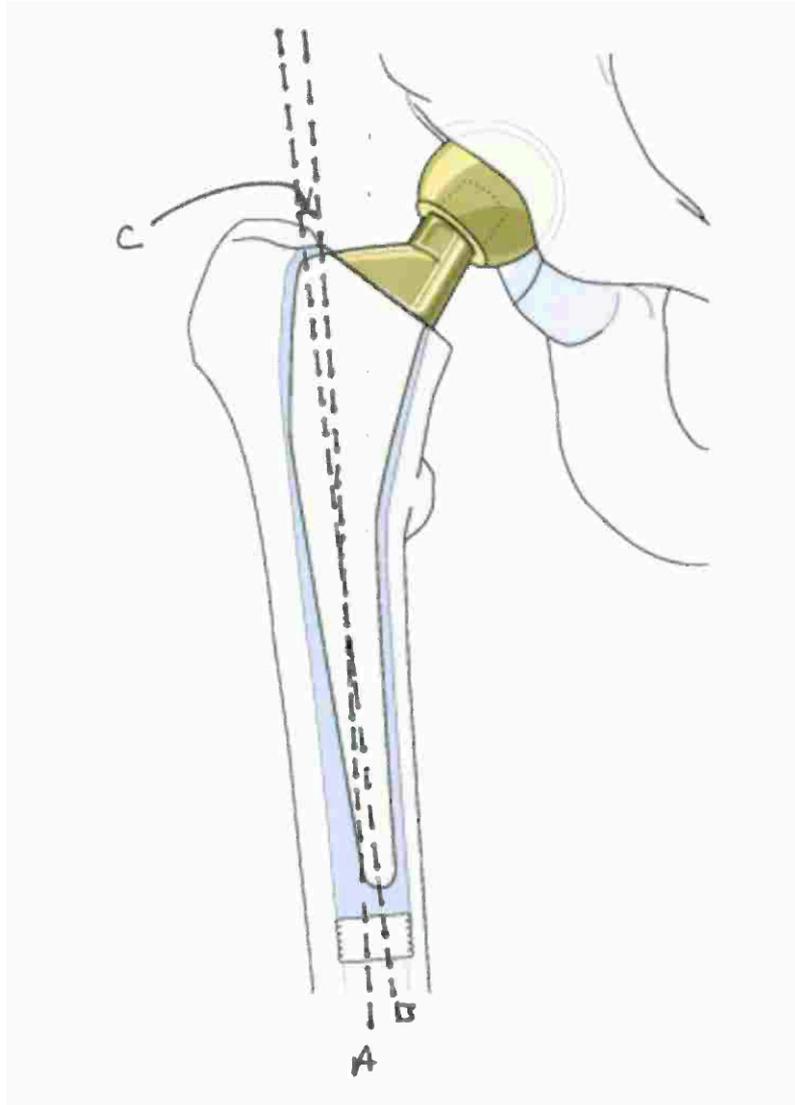
The pelvic reference line is drawn between the inferior borders of the acetabular teardrops. The femoral reference line is drawn between the lesser trochanters. The perpendicular distance is measured and compared for both sides.

(The bi-ischial line has been used as a pelvic reference but rotation of the film can make this less accurate⁴³¹)

2. Femoral stem alignment (McBride et al, 2011; Munuera et al, 1992)

The ideal femoral stem position is in neutral alignment with the longitudinal axis of the shaft⁴³¹

Varus malposition is associated with a failure rate of up to 46% at 16 years in cemented THA (there are no equivalent survival studies of cemented hemiarthroplasty)



A = longitudinal axis of the shaft
B = longitudinal axis of the femoral stem
C = degrees off neutral

3. Cement mantle grade (Barrack et al, 1992)

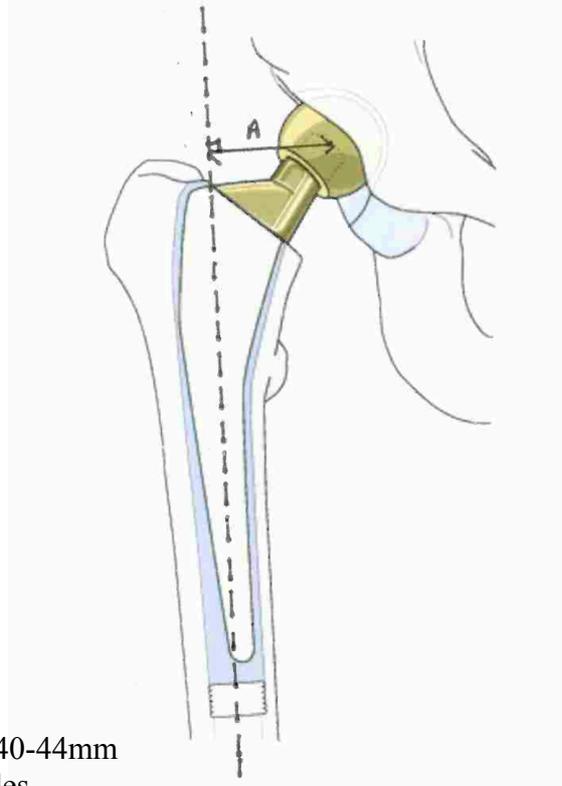
Barrack et al⁴³³ describe a cement mantle grading system;

Type A	Complete filling of the femoral canal with no distinguishable border between cement and bone ('white out')
Type B	Near complete filling with some demarcation between cement and bone Radiolucency at the cement-bone interface is <50%
Type C1	More than 50% lucency at the cement-bone interface
Type C2	The cement mantle is <1mm or the prosthesis is in contact with the bone
Type D	Gross deficiency or large voids

Poor cementation (Barrack C or D and a mantle less than 2mm are correlated with early failure)⁴³³

4. Femoral offset (Lecerf et al, 2009)

Femoral offset restoration is essential to improve function and longevity of hip arthroplasty⁴⁶⁶.



A = offset
Normal range = 40-44mm
Measure both sides

11.6 Chapter 7: Intra-operative blood loss calculator

Calculating intra-operative blood loss

The Haemoglobin balance method of calculating intra-operative blood loss (adapted from Gao et al.)

$$\mathbf{Hb_{loss\ total} = BV \times (Hb_i - Hb_e) \times 0.0001 + Hb_t}$$

$$\mathbf{V_{loss\ total} = 1000 \times Hb_{loss\ total} / Hb_i}$$

Hb_{loss total} (g): The loss of volume of Hb

Hb_i (g/L): The Hb value before surgery

Hb_e (g/L): The Hb value after surgery

Hb_t (g): The total volume of blood transfusion

V_{loss total} (ml): Volume of blood loss

BV (ml): The patient's blood volume before surgery.

I calculated BV using Nadler's equation (Nadler 1962);

$$\text{Male blood volume} = (0.3669 \times H3) + (0.03219 \times W) + 0.6041$$

$$\text{Female blood volume} = (0.3561 \times H3) + (0.03308 \times W) + 0.1833$$

Some hospitals did not record patient heights or weights, or did so inconsistently.

Once data collection was complete, I calculated average male and female heights and weights for the patients in the study, and imputed these gender averages for the Nadler formula for cases where the height and weight information was missing.

Gender was recorded for all cases.

11.7 Chapter 7: Sensitivity Analysis

Statistically significant results ($p < 0.05$) are shown in bold

DHS	As trained (per protocol)	As randomised (intention-to-train)
Acceptability of TAD χ^2	p<0.001	p=0.004
Position of lag screw in head χ^2	p<0.001	p=0.02
Learning curve	p=0.001	p=0.181
Procedure time (t-test)	p=0.271	p=<0.001
Radiation dose (t-test)	p=<0.001	p=0.038
Complication rate (χ^2)	p=0.251	p=0.084
Length of stay (t-test)	p=0.45	p=0.08
Mortality at 12months (χ^2)	p=0.87	p=0.033

HEMIARTHROPLASTY	As trained (per protocol)	As randomised (intention-to-train)
Acceptability of LLD (χ^2)	p=0.001	p=0.010
Stem alignment acceptability (χ^2)	p<0.001	p=0.001
Offset difference(χ^2)	p=0.342	p=0.578
Cement mantle quality	p=0.693	p=0.117
Procedure time (t-test)	p=0.07	p=0.008
PT learning curve	p=0.04	p=0.026
Blood loss (t-test)	p=0.21	p=0.241
Blood transfusion (χ^2)	p=0.003	p=0.001
Complication rate (χ^2)	p=0.036	p=0.193

Length of hospital stay (t-test)	p=0.053	p=0.915
Mortality at 12 months (χ^2)	p=0.25	p=0.493

ANKLE FRACTURE	As trained (per protocol)	As randomised (intention-to-train)
Acceptability of MCS (χ^2)	p=0.272	p=0.013
Learning curve analysis	p=0.203	p=0.431
Acceptability of LMD (χ^2)	p=<0.001	p=<0.001
Acceptability of MMD (χ^2)	p=0.009	p=0.001
Acceptability of TCS(χ^2)	p=<0.001	p=<0.001
Acceptability of TCA (χ^2)	p=0.528	p=0.235
Procedure Time (t-test)	p=0.66	p=0.921
PT learning curve	p=0.81	p=0.606
Radiation dose (t-test)	p=0.004	p=0.373
Learning curve	p=0.301	p=0.194

11.8 Chapter 8: Topic guide for interviews

Interview topic guide for CAD:TRAUMA participants

1. Demographic information about participant

- Current post including specialty and hospital
- Stage of training

2. Topics for discussion (not exhaustive)

- Challenges of modern day surgical training
- Experiences of cadaveric simulation as an adjunct to training
- Factors that make good surgeons
- Preparedness for operating in real-life as a junior trainee surgeon

Questions will be framed to encompass four key domains – knowledge, opinion, feeling and experience (1)

Example knowledge questions:

- 1) Do you know what the European Working Time Directive says about the hours trainee surgeons are allowed to work?
- 2) Do you know how many index procedure* performances and work-based assessments you need to have logged by CCT**

Example opinion questions:

- 1) What do you think about modern surgical training in general? In Trauma & Orthopaedic surgery? In your hospital?
- 2) What do you think about using simulation to augment training? Cadaveric simulation in particular?
- 3) What do you think makes a good surgeon?

Example feeling questions:

- 1) What do you feel is the best way to train surgeons?
- 2) Do you feel that the CST training has benefited? How so?

Example experience questions:

- 1) Can you tell me about any times that you were not able to get access to the training opportunities you needed in the operating theatre? If yes -Why do you think this happened? If no – How is your training organized to avoid this happening?
- 2) When you have been doing operations in the recent past, have you felt well prepared?

* Index procedures are the key operations as defined in the curriculum

** Certificate of Completion of Training – the endpoint of surgical training(1) Lichtman, M.

Qualitative research in education: A user's guide. 3rd Edition. London: Sage Publications; 2013



■ TRAUMA

Current provision of simulation in the UK and Republic of Ireland trauma and orthopaedic specialist training: a national survey

**H. K. James,
R. J. H. Gregory,
D. Tennent,
G. T. R. Pattison,
J. D. Fisher,
D. R. Griffin**

*Warwick Clinical Trials
Unit, Coventry, UK*

Aims

The primary aim of the survey was to map the current provision of simulation training within UK and Republic of Ireland (RoI) trauma and orthopaedic (T&O) specialist training programmes to inform future design of a simulation based-curriculum. The secondary aims were to characterize; the types of simulation offered to trainees by stage of training, the sources of funding for simulation, the barriers to providing simulation in training, and to measure current research activity assessing the educational impact of simulation.

Methods

The development of the survey was a collaborative effort between the authors and the British Orthopaedic Association Simulation Group. The survey items were embedded in the Performance and Opportunity Dashboard, which annually audits quality in training across several domains on behalf of the Speciality Advisory Committee (SAC). The survey was sent via email to the 30 training programme directors in March 2019. Data were retrieved and analyzed at the Warwick Clinical Trials Unit, UK.

Results

Overall, 28 of 30 programme directors completed the survey (93%). 82% of programmes had access to high-fidelity simulation facilities such as cadaveric laboratories. More than half (54%) had access to a non-technical skills simulation training. Less than half (43%) received centralized funding for simulation, a third relied on local funding such as the departmental budget, and there was a heavy reliance on industry sponsorship to partly or wholly fund simulation training (64%). Provision was higher in the mid-stages (ST3-5) compared to late-stages (ST6-8) of training, and was formally timetabled in 68% of programmes. There was no assessment of the impact of simulation training using objective behavioural measures or real-world clinical outcomes.

Conclusion

There is currently widespread, but variable, provision of simulation in T&O training in the UK and RoI, which is likely to expand further with the new curriculum. It is important that research activity into the impact of simulation training continues, to develop an evidence base to support investment in facilities and provision.

Cite this article: 2020;1-5:103–114.

Keywords: surgical training, simulation, education

Correspondence should be sent to
Hannah K James; email:

doi: 10.1302/2046-3758.15.BJO-
2020-0011

2020;1-5:103–114.

Introduction

Simulation has an increasingly important role in orthopaedic surgical training in a climate of reduced working hours,^{1,2} financial constraints^{3,4} and emphasis on patient safety.⁵ There is a move towards improving training efficiency and quality through

curriculum reform to a competency-based model⁶ and the improved surgical training initiative,⁷ both of which prominently feature simulation.

The main appeal of simulation as a surgical training adjunct is that it moves the early part of the surgical learning curve

away from patients⁸ into a controlled environment where competency can be measured and assured.⁹ It offers the potential for rapid ‘upskilling’, the expeditious attainment of surgical skill¹⁰ in a manner that is no longer achievable using the traditional master-apprentice model of ‘on-the-job’ training.

There are a wide variety of simulation training options available in orthopaedics, ranging from low fidelity bench-top box trainers^{11,12} and plastic bones^{13,14} for teaching basic orthopaedic skills, up to ultra-high fidelity simulation using human cadavers where an entire operation can be performed in a simulated operating theatre using real implants and instruments.^{15,16}

The potential benefits of simulation also include the acquisition of non-technical skills, which forms a significant part of the skill set of a competent surgeon, the importance of which is only recently gaining recognition.¹⁷ There is an emerging body of evidence that both technical and non-technical skills learnt in the simulated environment can be transferred to the operating theatre,^{18,19} with potential benefit to patients.

Accordingly, there is a move to increase simulation provision within trauma and orthopaedics (T&O) training. It is therefore timely and necessary to provide a comprehensive overview of current simulation provision in T&O training programmes.

Methods

Survey development. The primary aim of the survey was to map the current provision of simulation training within UK and the Republic of Ireland (RoI) T&O specialist training programmes to inform future design of a simulation based-curriculum. The survey questions were developed to describe, by programme and region; the resources available for simulation, sources of funding, timetabling of provision within training, the type of simulation offered at each stage of training, and research activity measuring the educational impact of simulation training.

Participants and survey administration. The survey questions were embedded in the 2019 performance and opportunity dashboard data collection cycle, which annually audits quality of training in T&O across several domains on behalf of the speciality advisory committee (SAC). The questions were flagged as being for research with a brief outline of the study aims, and completion of this section was not mandatory for successful submission to the dashboard. The survey, hosted by a commercial web platform (Survey Monkey, San Mateo, California, USA), was sent by email to the 30 training programme directors (TPDs) in the UK and RoI in March 2019. Three email reminders were sent to non-responders and the survey was closed in July 2019.

Data Analysis. Data were retrieved by the SAC quality assurance lead (RJHG) and sent to the Warwick Clinical Trial Unit for analysis (HKJ). A descriptive analysis was

undertaken, with all responses presented as counts and percentages. The results are presented visually as geo-thermal heat maps to show regional trends in provision ((Figs 1–6), respectively).

Results

Demographics. Overall, 28 out of 30 eligible TPDs completed the survey (93%). The geographical distribution of the 30 UK and ROI T&O training programmes are shown in Fig. 7. It should be noted that the geographical size of the programmes is independent of the number of trainees. The programme boundaries were determined by plotting the hospitals within a given rotation on Google maps (Google, Alphabet Inc, Mountain View, California, USA) and drawing boundaries around them. There are seven programmes in London, and considerable overlap between rotations among the London hospitals. A pragmatic, best-fit approach was therefore taken in determining the programme boundaries in London which allowed for a clear visual representation of the data.

Resources for simulation delivery. Over 96% of programmes had facilities for simulated clinical case management, such as clinical skills teaching rooms within hospitals. Just over half of programmes (54%) had access to non-technical skills simulation, delivering modules related to situational awareness, decision making, communication and teamwork. In terms of technical surgical skills teaching, 82% had access to cadaver-based surgical procedure simulation and 64% of programmes had access to arthroscopy simulation. Less than one-third (29%) of programmes had access to a simulated operating theatre environment. It is probable that the fifteen programmes offering cadaveric simulation without access to a simulated operating theatre are delivering this training within dissection laboratories affiliated with medical schools, rather than dedicated surgical training centres.

Four programmes (14%, three within London) offer access to simulation facilities to trainees out-of-hours on an ‘as required’ basis.

Simulation resources by programme were classified as very low, low, moderate, high and very high according to the facilities available, and are shown in Fig. 1 with descriptors. There was no clear geographical pattern to provision of facilities for simulation, and there was considerable variation regionally. Training (on-line, on-paper, or face-to-face) was available for the simulation trainers in five regions (shown in Fig. 1 by the dotted shading), and this provision bore no relation to the overall quality of simulation facilities within the training programme. The opportunity to be involved in simulation training was made available to all trainers in just under half (46%) of programmes.

Sources of funding. Less than half (43%) of training programmes received centralised funding for simulation provision from Health Education England and/or the

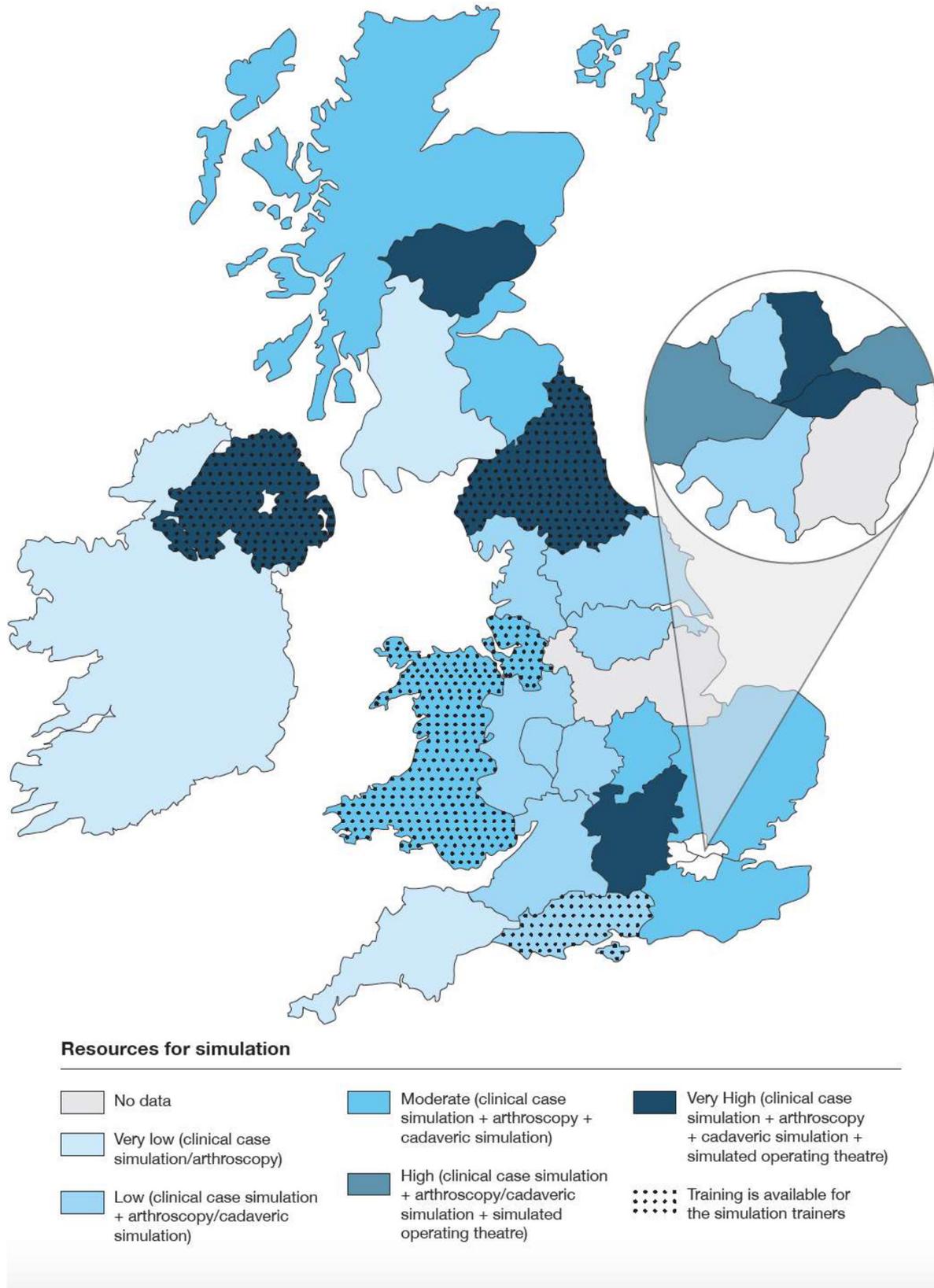


Fig. 1

Resources for simulation training by programme.

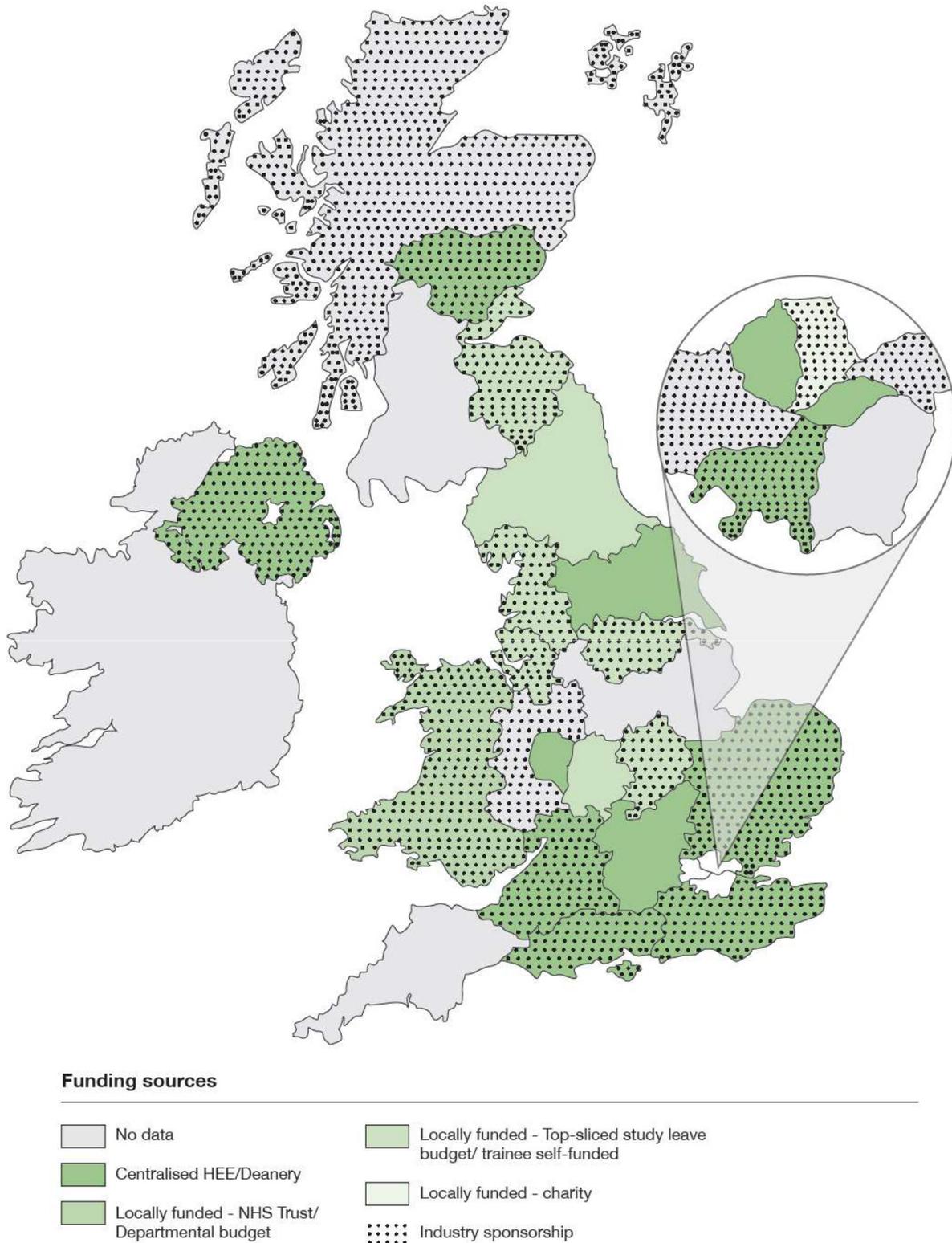


Fig. 2

Sources of funding for simulation training.

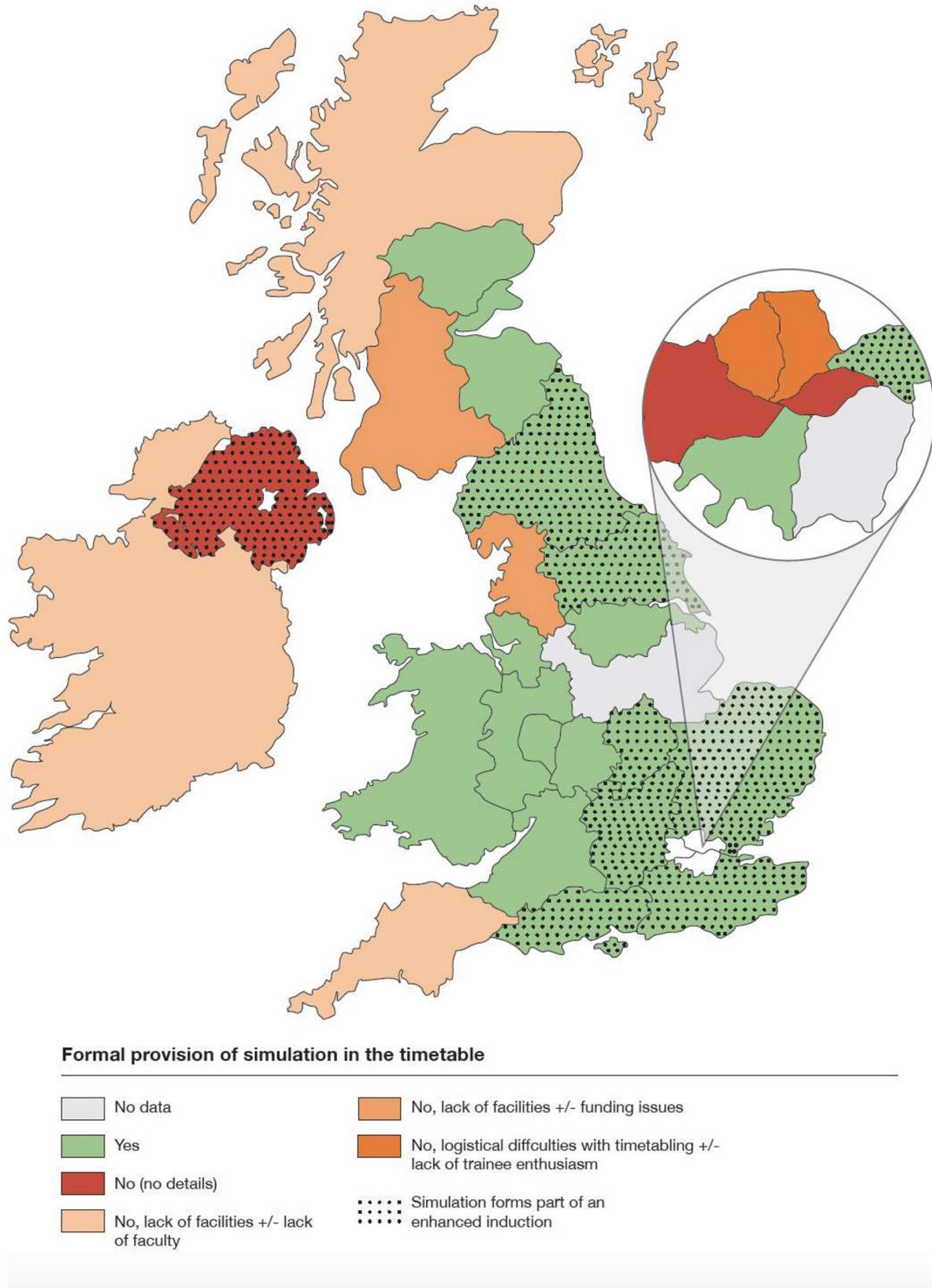


Fig. 3
Provision within training timetable and barriers to delivery.

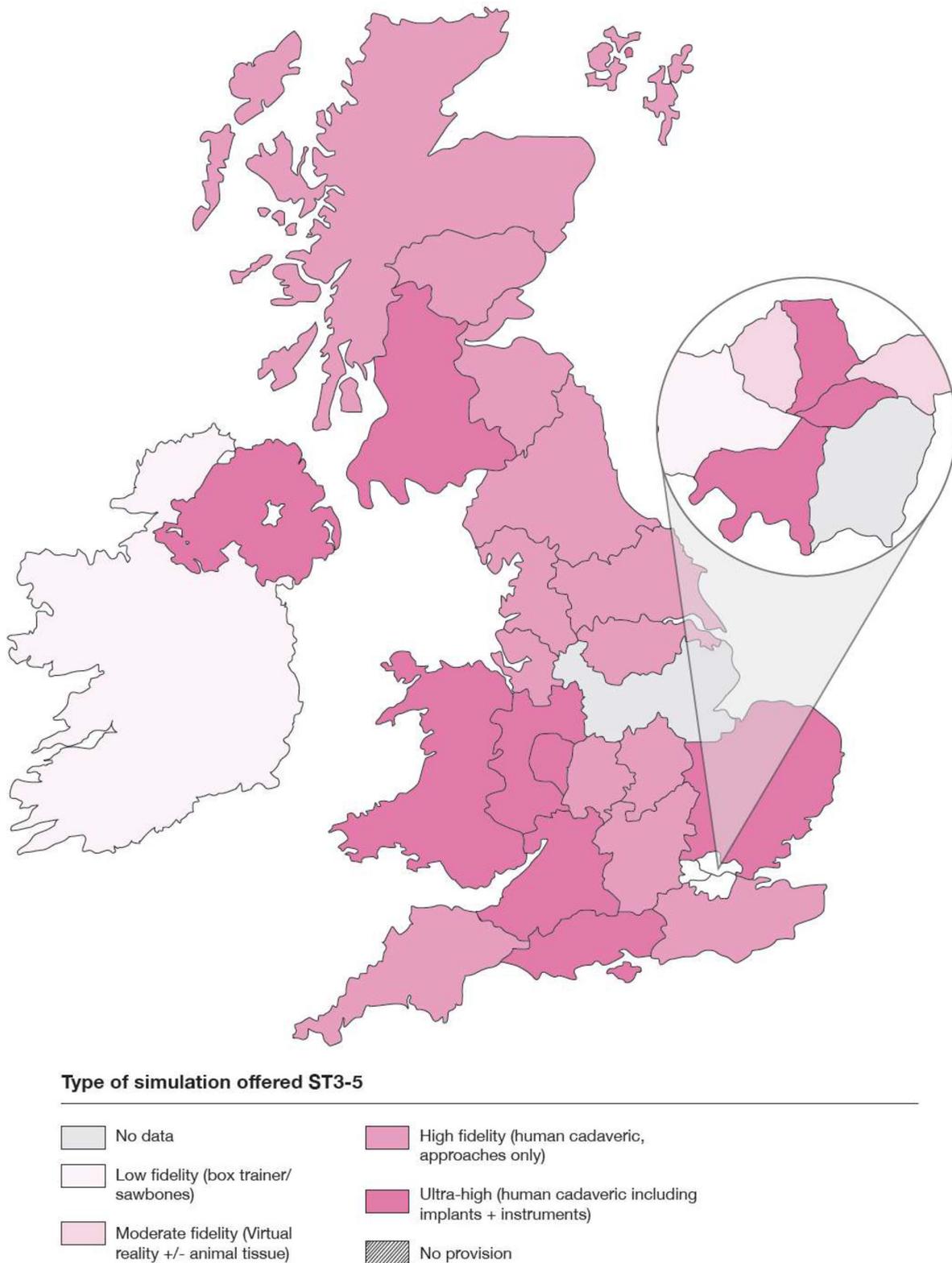


Fig. 4

Provision by programme at the mid-stage of training (ST3-5).

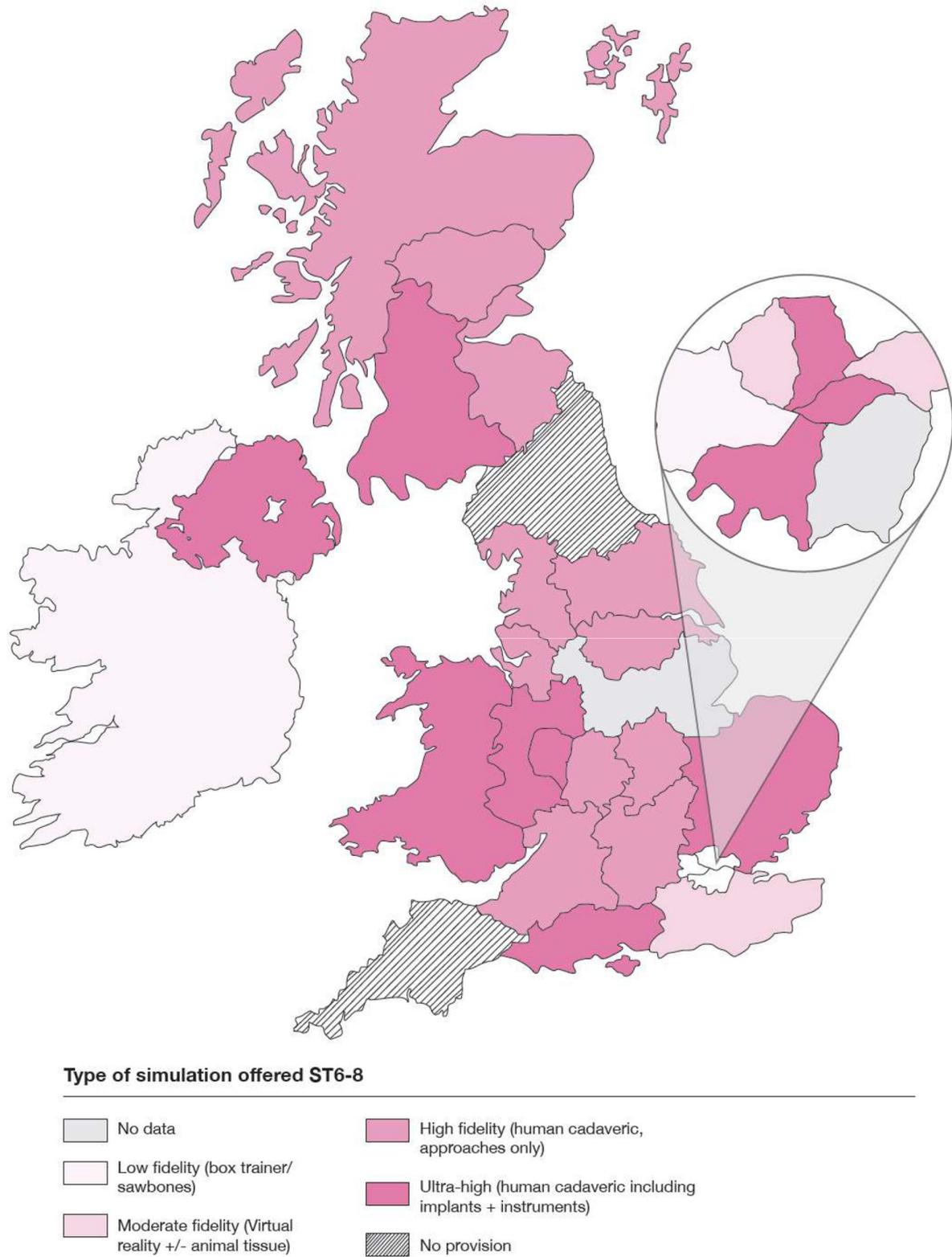


Fig. 5

Provision by programme at the late stage of training (ST6-8).

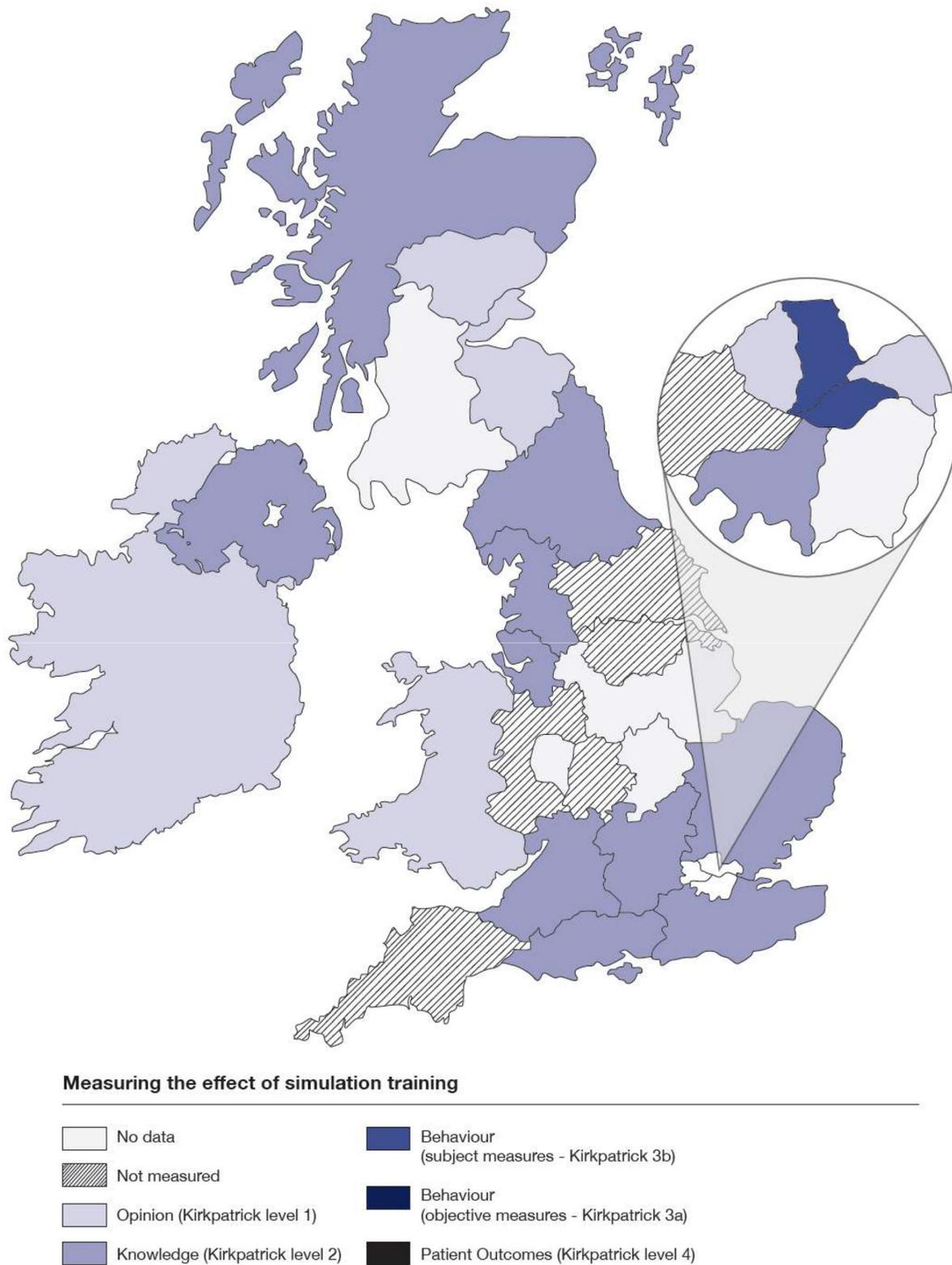


Fig. 6

Research activity to measure the educational impact of simulation provision in training.

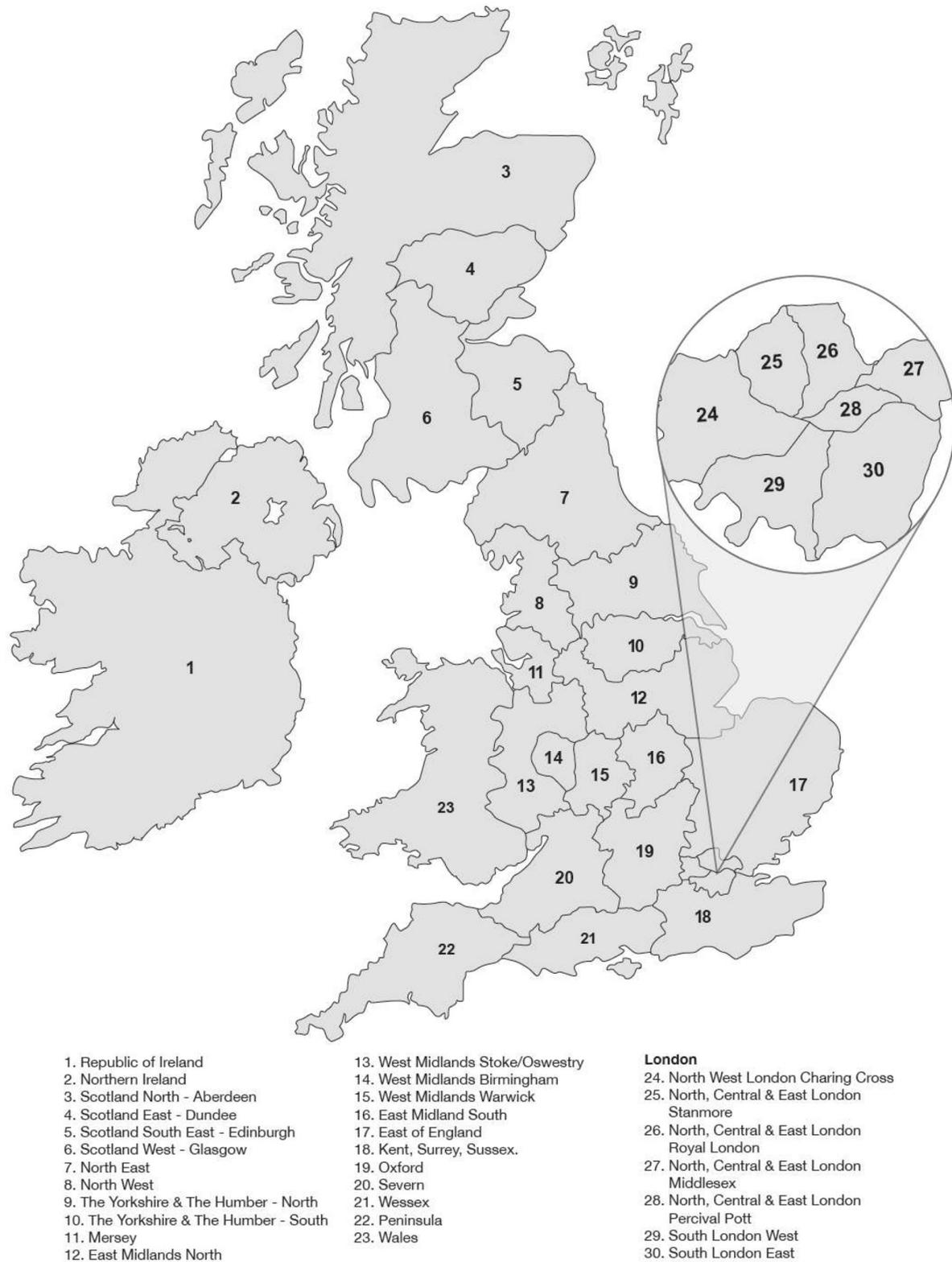


Fig. 7

Geographical distribution of training programmes in the UK and RoI.

postgraduate deanery (Fig. 2). One-third (32%) of programmes funded simulation provision locally, using either NHS trust/departmental budget (two programmes), top-slicing trainee study budgets or trainee self-funding (six programmes) or via a charitable foundation (one programme). Of note, two-thirds (64%) of programmes received industry or other commercial sponsorship for the provision of simulation. There was no relationship between lack of centralized funding and reliance on industry sponsorship, and quantitative estimates of funding were not sought in the survey. The programmes with centralized funding support for simulation were mainly clustered in the south of England (Fig. 2). No funding information was given for six programmes.

Barriers to provision. The TPDs were asked about obstacles to provision of simulation within their respective programmes (Fig. 3). Overall, 40% of TPDs reported barriers to delivery of simulation; six programmes reported lack of facilities, two had funding issues, two reported logistical difficulties related to timetabling, one reported lack of available faculty, and one reported lack of trainee enthusiasm.

Formal provision of simulation within the timetable. Simulation was formally timetabled in 19 (68%) of training programmes, and was used as part of an enhanced induction programme in nine (32%). There was a clear relationship between the provision of a simulation enhanced induction programme and formal timetabling of simulation, with only one programme offering the former without the latter. Simulation was more likely to be formally provided in the timetable in England and Wales than in Scotland and Ireland/Northern Ireland (Fig. 3).

Type of simulation offered by stage of training. Provision of simulation by type and stage was measured, with mid (ST3-5) and late (ST6-8) specialist training considered separately. Low fidelity simulation was defined as box trainer or Sawbones (Sawbones Europe, Malmö, Sweden) or equivalent. Moderate fidelity was defined as virtual reality (VR) or animal tissue. High fidelity was defined as human cadaveric simulation, involving surgical approaches only, and ultra-high fidelity was the addition of implants, instruments and the ability to perform the entire procedure on the cadaver. Type of simulation offered by stage of training by programme is shown in Figs 4 and 5. As would be expected, the degree of fidelity offered within programmes broadly correlates with the facilities available for simulation (Fig. 1.) There is a trend towards less provision at the later stages of training, with two programmes offering no simulation provision at ST6-8, and three programmes offering high and ultra-high fidelity simulation to ST3-5 only.

Measuring the impact of simulation. Research activity measuring the educational impact of simulation provision was stratified according to Kirkpatrick's hierarchy,²⁰ which is a widely accepted framework for classifying

the educational outcomes of a training intervention. Six programmes (21%) did not measure the impact of their simulation training. Six programmes (21%) measured the effect of simulation training on trainee opinion (Kirkpatrick level 1), which typically involves using pre- and post-training questionnaires to assess change in subjective metrics such as confidence. In all, 11 programmes (40%) assessed simulation provision using changes in trainee knowledge (Kirkpatrick level 2), through either bespoke post-training knowledge testing or the annual UKITE exercise. Neither of these methods assess technical or clinical skill improvement from simulation or transfer to the workplace. Two programmes in London (Fig. 6) report measuring the impact of simulation training using subjective behavioural measures (Kirkpatrick 3b). It is striking that there is no assessment of the impact of simulation within training programmes using the highest level of educational impact metrics - objective behavioural measures (Kirkpatrick level 3a) or patient outcomes (Kirkpatrick level 4).

Discussion

Orthopaedic educators face the considerable challenge of continuing to train high-quality surgeons amid increasing clinical, financial, regulatory and time pressures on training.²¹ The traditional master-apprentice model of surgical training, with its reliance on the two central tenets of high volume of case exposure and a sustained mentor-mentee relationship,²² has been rendered obsolete in the modern surgical healthcare environment.⁸ This is due largely to increased service demands at the expense of training,²³ shift based working patterns, and short training rotations.

There is a growing evidence base supporting the use of simulation as an adjunct to training, showing that the learning curve can be advanced away from patients,¹⁰ with inherent safety advantages,²⁴ and that skills learnt in the simulated environment can transfer to the operating theatre.¹⁹ There have been many studies demonstrating the face²⁵ and construct^{26,27} validity, feasibility²⁸ and educational impact^{29,30} of various orthopaedic simulators for both open and arthroscopic surgery, ranging from low-fidelity, low-cost box-trainer type models through to ultra-high-fidelity cadaveric simulation.

This study is the first to map simulation in T&O training, and describe its current provision status on a national scale. The results of the study show that overall, simulation provision is highly variable across the 30 T&O training programmes of the UK and RoI. The availability of resources varies widely (Fig. 1), and is likely to be at least partly influenced by the geographical and financial relationship of training programmes with University medical schools, where cadaveric wet-laboratory facilities are generally situated.

The funding landscape for simulation provision is complex and not explored in detail in this study. There was a tendency for simulation provision in the southern half of England to be described as centrally funded, the reason for which is unclear, and a widespread reliance on industry sponsorship (64%) was seen, seemingly independent of centralized funding status. Of note, four programmes reported that trainees were required to self-fund simulation training. This raises obvious ethical concerns around equity of access to training opportunity, as some trainees may be disadvantaged through being unable to pay to access simulation training. Six studies reported funding simulation training by top-slicing trainees' study budget, which although fairer than asking trainees to self-fund, erodes their already limited study budget and may jeopardise access to other valuable training opportunities such as attendance at courses and conferences.

Barriers to formally integrating simulation in the timetable were explored. Barriers cited included; lack of facilities (n = 6), lack of faculty (n = 1), funding issues (n = 2), logistical issues with timetabling (n = 2) and, surprisingly, lack of trainee enthusiasm (n = 1). The majority of programmes (68%) have formally timetabled simulation provision, which is important for logistical considerations when planning training, and also assures the trainees of protected 'bleep-free' teaching time. Formal timetabling also removes the need for a separate study leave application process or annual leave usage to attend simulation training, and attendance rates can be monitored.

Type of simulation offered by stage of training and region is shown in Figs 4 and 5. A disparity is seen between available facilities and reported provision in Glasgow, where ultra-high fidelity training is provided at both the mid- and late-stages of training but the resources for simulation is given as very low with no cadaveric facilities. One explanation is that cadaveric training is outsourced to another region, which we did not address in the survey. Less simulation provision was generally available during late stage training (ST6-8). This may be explained by the potentially greatest gains being obtained from simulation at an earlier stage of training, where the learning curve is steeper.

A limitation of this study is that we have focussed on establishing facts about provision at the expense of a more nuanced understanding of TPD opinion on the role of simulation in their training programmes. We have also not included trainees in this study, who as the direct beneficiaries of simulation training, should be central to discussions around provision. We obtained a very high response rate of 93% but the survey cannot claim to be comprehensive as there are two training programmes for which we have no data. Provision of simulation is only one of several quality indicators relevant to assessing a training programme, and we make no inferences as to

the quality of individual training programmes based on these results.

Despite the obvious appeal of simulation as a partial solution to the challenges of the modern surgical training environment, there is no evidence to date that simulation training in orthopaedics benefits patients (Kirkpatrick level 4 evidence). Orthopaedics lags behind general surgery in its efforts to measure the educational impact of simulation training, in a systematic review¹⁸ of skill transfer to the operating theatre after simulation training only one³¹ of 34 studies was from orthopaedics. Research efforts to measure so called 'transfer validity' have been complicated by a lack of appropriately validated objective outcome measures,³² and are largely restricted to the research setting. Only two studies^{31,33} have shown evidence of transfer validity following simulation training in orthopaedics, both of these involve diagnostic knee arthroscopy. Arthroscopic procedures lend themselves more easily to objective measurement of skill transfer, as motion analysis can be used to objectively measure performance in the simulated environment, and subsequently the operating theatre. The measurement of transfer validity of open procedures is considerably harder.³²

Simulation delivery can be costly, cadaveric training especially so.³⁴ Until such a time that there is a high-quality evidence base showing that simulation training in T&O improves technical and non-technical skills that translate into the workplace for the benefit of patients, it is unlikely that simulation is going to be mandated for training or comprehensively funded by HEE.

The survey results show that despite some challenges there is currently widespread simulation provision in training, much of it using sophisticated techniques such as cadaveric simulation (Figs 4 and 5). The use of innovative funding streams, and widespread efforts, albeit imperfect, to measure the educational benefit of simulation suggests a high level of enthusiasm for the delivery of simulation in T&O. These results reveal a promising foundation for the future of simulation delivery in T&O training, which we anticipate will continue to grow further with the implementation of the new curriculum next year. It is important that research activity continues keeps pace with the anticipated expansion of simulation provision, and can inform future developments in an evidence based manner.

Twitter

Follow H. James @hannah_ortho

References

1. Pickersgill T. The European working time directive for doctors in training. *BMJ*. 2001;323(7324):1266.
2. House J. Calling time on doctors' working hours. *The Lancet*. 2009;373(9680):2011–2012.

3. The state of medical education and practice in the UK. London: General Medical Council, 2016.
4. **Department of Health - Third Report.** *House of commons, editor.* London: The Stationary Office, 2008.
5. **Gaskell N, Hinton R, Page T, Elvins T, Malin A.** Putting an end to Black Wednesday: improving patient safety by achieving comprehensive trust induction and mandatory training by day 1. *Clin Med.* 2016;16(2):124–128.
6. **Bowditch MG.** Editor curriculum update. British orthopaedic association annual Congress. Liverpool, 2019.
7. **Improving Surgical Training.** Proposal for a pilot surgical training programme. 2015.
8. **Reznick RK, MacRae H.** Teaching surgical skills — changes in the wind. *N Engl J Med.* 2006;355(25):2664–2669.
9. **MacCaskie A, Kenny D, Deshmukh S.** How can surgical training benefit from theories of skilled motor development. *musical skill acquisition and performance psychology? The Medical Journal of Australia.* 2011;194(9).
10. **Curran I.** A framework for technology enhanced learning. London: Department of Health, 2011.
11. **Hohn EA, Brooks AG, Leasure J, et al.** Development of a surgical skills curriculum for the training and assessment of manual skills in orthopedic surgical residents. *J Surg Educ.* 2015;72(1):47–52.
12. **Ruder JA, Turvey B, Hsu JR, Scannell BP.** Effectiveness of a low-cost drilling module in orthopaedic surgical simulation. *J Surg Educ.* 2017;74(3):471–476.
13. **LeBlanc J, Hutchison C, Hu Y, Donnon T.** A comparison of orthopaedic resident performance on surgical fixation of an ulnar fracture using virtual reality and synthetic models. *The Journal of Bone & Joint Surgery.* 2013;95(9):e60–65.
14. **Leong JJH, Leff DR, Das A, et al.** Validation of orthopaedic bench models for trauma surgery. *J Bone Joint Surg Br.* 2008;90-B(7):958–965.
15. **Gottschalk MB, Yoon ST, Park DK, Rhee JM, Mitchell PM.** Surgical training using three-dimensional simulation in placement of cervical lateral mass screws: a blinded randomized control trial. *The Spine Journal.* 2015;15(1):168–175.
16. **Camp CL, Krych AJ, Stuart MJ, et al.** Improving resident performance in knee arthroscopy: a prospective value assessment of simulators and cadaveric skills laboratories. *Journal of Bone and Joint Surgery - American Volume.* 2016;98(3):220–225.
17. **Agha RA, Fowler AJ, Sevdalis N.** The role of non-technical skills in surgery. *Annals of Medicine and Surgery.* 2015;4(4):422–427.
18. **Dawe SR, Pena GN, Windsor JA, et al.** Systematic review of skills transfer after surgical simulation-based training. *Br J Surg.* 2014;101(9):1063–1076.
19. **Sturm LP, Windsor JA, Cosman PH, et al.** A systematic review of skills transfer after surgical simulation training. *Ann Surg.* 2008;248(2):166–179.
20. **Kirkpatrick D.** Techniques for evaluating training programmes. *Journal of American Society for Training and Development.* 1959;13(11):11–12.
21. **Karam MD, Pedowitz RA, Natividad H, Murray J, Marsh JL.** Current and future use of surgical skills training laboratories in orthopaedic resident education: a national survey. *J Bone Joint Surg Am.* 2013;95(1):e4.
22. **Walter AJ.** Surgical education for the twenty-first century: beyond the apprentice model. *Obstet Gynecol Clin North Am.* 2006;33(2):233–236.
23. **James HK, Fisher JG, Griffin DR, Pattison G.** If we can't get to theatre, we can't learn to operate: A qualitative study of factors influencing core trainee access to the operating theatre in Trauma & Orthopaedics. *Ann R Coll Surg Engl.* In Press. 2019.
24. **Scott DJ, Safety P.** Patient safety, competency, and the future of surgical simulation. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare.* 2006;1(3):164–170.
25. **Tuijthof GJM, van Sterkenburg MN, Sierevelt IN, et al.** First validation of the PASSPORT training environment for arthroscopic skills. knee surgery, sports Traumatology. *Arthroscopy.* 2010;18(2):218–224.
26. **Angelo RL, Ryu RKN, Pedowitz RA, Gallagher AG.** Metric development for an arthroscopic Bankart procedure: assessment of face and content validity. *Arthroscopy: The Journal of Arthroscopic & Related Surgery.* 2015;31(8):1430–1440.
27. **Colaco HB, Hughes K, Pearse E, Arander M, Tennent D.** Construct validity, assessment of the learning curve, and experience of using a low-cost arthroscopic surgical simulator. *J Surg Educ.* 2017;74(1):47–54.
28. **Coughlin RP, Pauyo T, Sutton JC, Coughlin LP, Bergeron SG.** A validated orthopaedic surgical simulation model for training and evaluation of basic arthroscopic skills. *J Bone Joint Surg Am.* 2015;97(17):1465–1471.
29. **Cannon WD, Nicandri GT, Reinig K, Mevis H, Wittstein J.** Evaluation of skill level between trainees and community orthopaedic surgeons using a virtual reality arthroscopic knee simulator. *J Bone Joint Surg Am.* 2014;96(7):e57.
30. **Rebolledo BJ, Hammann-Scala J, Leali A, Ranawat AS.** Arthroscopy skills development with a surgical simulator: a comparative study in orthopaedic surgery residents. *The American journal of sports medicine.* 2015;43(6):1526–1529.
31. **Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL.** Transferring simulated arthroscopic skills to the operating theatre. *J Bone Joint Surg Br.* 2008;90-B(4):494–499.
32. **James HK.** Measuring the educational impact of simulation training in Trauma & Orthopaedics. *Journal of Trauma & Orthopaedics.* In Press. 2019.
33. **Garfield Roberts P, Alvand A, Gallieri M, Hargrove C, Rees J.** Objectively assessing intraoperative arthroscopic skills performance and the transfer of simulation training in knee arthroscopy: a randomized controlled trial. *Arthroscopy: The Journal of Arthroscopic & Related Surgery.* 2019;35(4):1197–1209.
34. **Gilbody J, Prasthofer AW, Ho K, Costa ML.** The use and effectiveness of cadaveric workshops in higher surgical training: a systematic review. *Annals.* 2011;93(5):347–352.

Author information:

- H. K. James
- D. R. Griffin
Clinical Trials Unit, Warwick Medical School, Coventry, UK; Department of Trauma & Orthopaedic Surgery, University Hospitals Coventry & Warwickshire, Coventry, UK.
- R. J. H. Gregory, Department of Trauma & Orthopaedic Surgery, University Hospital of North Durham, Durham, UK.
- D. Tennent, Department of Orthopaedics, St. George's NHS Trust, London, UK.
- G. T. R. Pattison, Department of Trauma & Orthopaedic Surgery, University Hospitals Coventry & Warwickshire, Coventry, UK.
- J. D. Fisher, Clinical Trials Unit, Warwick Medical School, Coventry, UK.

Funding statement:

- H. K. James holds a Versus Arthritis Educational Research Fellowship (grant number 20485).

© 2020 Author(s) et al. This article is distributed under the terms of the Creative Commons Attributions (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original author and source are credited.

How Does Cadaveric Simulation Influence Learning in Orthopedic Residents?

Hannah K. James, MBChB (Hons), MRCS, MMedEd, ^{*,†}

Giles T.R. Pattison, MBBS FRCSEd (Tr&Orth) FHEA MSc FFSTEd, [†]

Damian R. Griffin, MBM BCH (Oxon) MA MPhil (Cantab) FRCS (Tr&Orth), ^{*,†} and Joanne D. Fisher, PhD^{*}

^{*}Clinical Trials Unit, Warwick Medical School, Coventry, United Kingdom; and [†]Department of Trauma and Orthopaedic Surgery, University Hospitals Coventry and Warwickshire, Coventry, United Kingdom

OBJECTIVE: The objectives of this study were to understand how cadaveric simulation impacts learning in orthopedic residents, why it is a useful training tool, and how skills learnt in the simulated environment translate into the workplace.

DESIGN: This is a qualitative research study using in-depth, semistructured interviews with orthopedic residents who underwent an intensive cadaveric simulation training course.

SETTING: The study was conducted at the University Hospital Coventry & Warwickshire, a tertiary care center with integrated cadaveric training laboratory in England, United Kingdom.

PARTICIPANTS: Orthopedic surgery residents in the intervention group of a randomized controlled trial comparing intensive cadaveric simulation training with standard “on the job” training were invited to participate. Eleven of 14 eligible residents were interviewed (PGY 3-6, 8 male and 3 female).

RESULTS: Learning from cadaveric simulation can be broadly categorized into intrinsic, surgeon-driven factors, and extrinsic environmental factors. Intrinsic factors include participant ability to “buy-in” to the simulation exercise, willingness to push one’s own learning boundaries in a “safe space” and take out on resident experience and self-reported confidence, with the greatest learning gains seen at around the PGY4 stage in individuals who reported low preintervention operative confidence. Extrinsic factors

included; the opportunity to perform operations in their entirety without external pressures or attending “take-over,” leading to subjective improvement in participant operative fluency and confidence. The intensive supervision of subspecialist attending surgeons giving real-time performance feedback, tips and tricks, and the opportunity to practice unusual approaches was highly valued by participants, as was paired learning with alternating roles as primary surgeon/assistant and multidisciplinary involvement of scrub-staff and radiographers. Cadaveric simulation added educational value beyond that obtained in low-fidelity simulation training by “stirring into practice” and “becoming through doing.” In providing ultrarealistic representation of the space, ritualism, and costuming of the operating theater, cadaveric simulation training also enabled the development of a range of nontechnical skills and sociocultural “nontechnical” lessons of surgery.

CONCLUSIONS: Cadaveric simulation enhances learning in both technical and nontechnical skills in junior orthopedic residents within a single training package. Direct transfer of skills learnt in the simulation training to the real-world operating theater, with consequent patient benefit, was reported. Cadaveric simulation in the UK training system of orthopedics may be of greatest utility at around the PGY 4 stage, at which point operative fluency, independence, and confidence can be rapidly improved in the cadaveric laboratory, to enable the attainment of competence in index trauma operations. (J Surg Ed 000:1–12. © 2019 The Authors. Published by Elsevier Inc. on behalf of Association of Program Directors in Surgery. This is an open access article under the CC BY-NC-ND license.

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

ABBREVIATIONS: PGY, postgraduate year CST, cadaveric simulation training EWTD, European Working Time Directive ORIF, open reduction internal fixation SHO,

Funding: TBC. HJ holds an educational research fellowship with Versus Arthritis (formerly Arthritis Research UK) grant number 20485. Material costs for the course were supported by Industry (DePuy Synthes UK, Leeds, England) and a grant from Health Education England (West Midlands Deanery).

Correspondence: Inquiries to Hannah K. James, Clinical Sciences Research Laboratories, University Hospital Coventry and Warwickshire, Clifford Bridge Road, Coventry, CV2 2DX, United Kingdom; e-mail: [REDACTED]

senior house officer CCT certificate of completion of training

KEY WORDS: high-fidelity simulation, surgical education, orthopedic residency, preparing for practice, cadaveric simulation training

COMPETENCIES: patient care, medical knowledge, practice-based learning and professionalism, professionalism

INTRODUCTION

The movement to incorporate simulation into postgraduate surgical training is rapidly gaining momentum, in response to perceived threats to the quality of training.¹⁻⁶ In the United Kingdom, a 2-year internship, undertaken during postgraduate year (PGY) 1 to 2, is followed by competitive entry into basic surgical training, which begins at PGY3 level. Basic surgical training comprises PGY 3 and 4, at which point a further competitive entry procedure occurs for progression into higher surgical training, which runs from PGY 5 to 10. On completion of higher surgical training, a surgeon is awarded a certificate of completion of training and can begin independent consultant practice at attending-equivalent level.

Surgical training in the United Kingdom is still largely based on the traditional Halstedian master-apprentice model,⁷ the success of which relies on an environment of long working hours, volume of exposure,⁸ unstructured training progression, and the maintenance of long-term working relationships with senior surgeons to foster constructive mentorship.⁹ However, in the current surgical training climate of shift-based work patterns,¹⁰ legally mandated reduced working hours,^{11,12} and a move to expediting surgical training,^{13,14} new models of training need to be considered to ensure the continued production of appropriately skilled practitioners.

Simulation offers a potential solution to some of these challenges by enabling rapid skill acquisition and progression to competency,³ within a safe, structured, and controlled environment remote from patients. Other safety-critical professions, most notably aviation,¹⁵ have long used simulation at early career stages to enable trainee pilots to achieve, and demonstrate, competency in the necessary skills before being allowed to take the controls of a real aeroplane. The logical extension of this argument is consideration of why, in surgical training, it is still considered acceptable to allow trainee surgeons to perform their first attempt at a whole operation on a real live patient, with the inherent risks of making mistakes as an inevitable part of learning, with potentially harmful patient consequences. The steep part of the learning curve, from novice to competent (Fig. 1) could be moved away from the patient to the simulator, and

only once a trainee has achieved a defined level of competence will he or she be permitted to begin operating on real patients.

Cadaveric simulation (i.e., using deceased, preserved, or fresh donated human bodies) is a potentially promising modality for training as it offers what is arguably the most realistic representation of human anatomy,¹⁶ with realistic tissue handling characteristics and haptic “touch-real” feedback,¹⁷ which cannot be easily replicated by other means.¹⁸

There is a drive to embed cadaveric simulation in surgical training in the United Kingdom, as part of widespread government-led efforts to modernize and improve efficiency of training. There has been considerable investment in facilities and expansion of provision of cadaveric simulation in recent years.

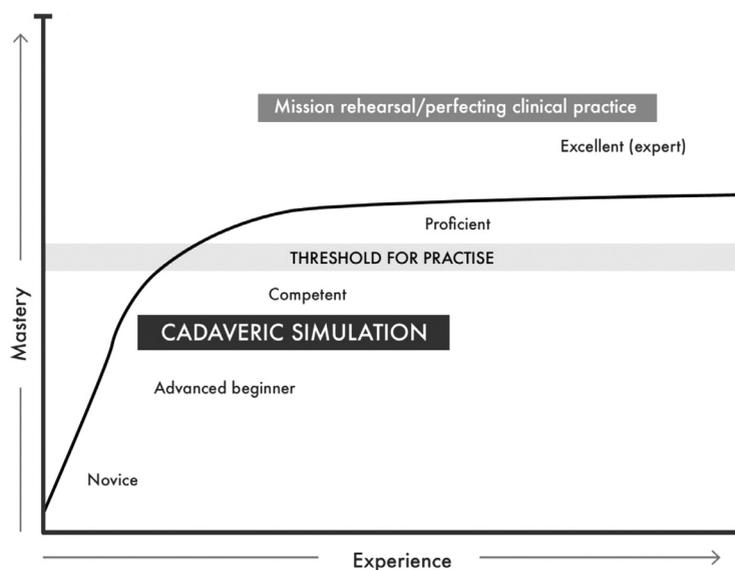
The literature on the qualitative impact of simulation training for surgeons is limited, and more widely, surgical simulation research is generally undertaken from a quantitative, reductionist perspective.¹⁹ This is because the measures of effectiveness and efficiency of training relate primarily to the mastery of technical skills and confidence gains, including “downstream translational effects on patient outcomes and patient care practices.”¹⁹ A criticism of restricting simulation-based research to purely outcome and effectiveness-based studies is that our understanding of the benefits of simulation-based education remains unidimensional, and that where rich contextualized detail of its impact is missing from the evidence base, explanation of phenomena that could be used to improve the educational value of future simulation training activity is not possible.

The primary objective of this study is to seek a rich understanding of the experiences of surgeons undertaking cadaveric simulation training (CST). The secondary objectives are to describe learner perspectives of the value of CST as a training tool, and to understand how skills learnt in a simulated environment are translated into the workplace from the perspective of the lived experience of the participants.

METHODS

Setting and Participants

This qualitative research study was carried out at the University Hospital Coventry & Warwickshire, a large tertiary care center in the West Midlands of England, United Kingdom. Within the hospital, there is a leading surgical training suite, hosting an active program of regional, national, and international cadaveric training courses. Ethical approval was granted for this research (Biomedical and Scientific Research Ethical Committee REGO-2014-718). This study was embedded within a



Adapted from Curran I. A Framework for Technology Enhanced Learning. London: Department of Health; 2011

FIGURE 1. Schematic learning curve with Dreyfus stages
Adapted from ref. [3].

randomized controlled trial comparing the impact of an intensive CST course on objectively measured real-world operative performance versus no additional training beyond the current standard “on the job” training.

Trial participants were a cohort of PGY3-6 orthopedic residents, and those in the intervention group ($n = 14$) were invited by e-mail to participate in the qualitative arm of the study. First and second e-mail reminders were sent to initial nonresponders. Eleven participants were subsequently interviewed. Of the 3 trial participants who were not interviewed, 1 declined to be interviewed, 1 had emigrated abroad into a nonclinical job role, and 1 participant did not respond to attempts at contact.

The mean age of participants was 28 years (range 26-31), 8 participants were male and 3 female. There was 1 in PGY3, 5 in PGY 4, and 5 in PGY5. Four participants had previously experienced CST in any capacity, and 7 were CST-naïve.

All interviews were conducted 6 months after completion of the CST intervention.

The CAD:TRAUMA Cadaveric Training Course

As part of the randomized trial, participants underwent an intensive, 2-day CST course. Eight waist-to-toe tip fresh-frozen cadavers were set-up supine on operating tables and 2 identical circuits of 4 stations were run in parallel. Participants were paired, each pair worked on 1 cadaver performing 1 operation (as first surgeon and

assistant) under the supervision of attending faculty, before rotating to the next station after a predetermined time. Over the 2 days, each participant acted as both first surgeon and assistant to their partner for the 4 procedures, and hence each participant was exposed to 8 procedures in their entirety; 4 as first surgeon and 4 as assistant. The 4 procedures were (1) dynamic hip screw fixation, (2) cemented hemiarthroplasty for fractured neck of femur, (3) plate and screw fixation for fractured ankle (open reduction internal fixation—“ORIF”), and (4) 4-compartment lower limb fasciotomy.

These procedures were chosen as they are curriculum-defined “index” procedures in which competency is expected to be achieved for progression from PGY4 to PGY5 in orthopedic residency (which within the UK system represents a significant transition, with a considerable increase in operative and decision-making responsibility from PGY4 level as a “Senior House Officer”/“SHO” to PGY5 level and onward as a “Specialist Registrar”). Great effort was made to authentically recreate the operating theater environment to ensure the highest possible environmental fidelity of the simulation; each cadaver was fully draped and participants wore surgical scrubs, masks, gloves, hats, and gowns. Each station had operating theater lights in use and fully stocked instrument trays with an array of implants (and cement for the hemiarthroplasty station). Scrub nurses and radiographers participated as ancillary faculty, and there were 2 radiographers with “mini-C” arm x-ray machines to check intra/postoperative implant position. The

preprocedure preparation that would be undertaken as routine in real-life surgery, such as patient positioning, skin preparation, draping, and incision site marking were also carried out before each case. The participants were expected to “scrub up” for each case and to respect the sterile operative field as they would in real life. The result was 8 highly realistic, equipped and staffed operating theaters running in tandem within the surgical training suite.

The consumable costs of the course were funded by a research grant, and used commercially procured fresh-frozen human cadaveric material (Science Care Ltd, Phoenix, Arizona) and industry donated implants and instruments (DePuY Synthes, Raynham, MA). The approximate cost of the course was \$2500/delegate. The attending faculty generously donated their “continuing professional development” time to the course, so there was no extra faculty cost incurred.

Data Collection

Interviews took place 6 months after conclusion of the CST intervention. This was felt to be the optimum post-training interval, with an important balance to be struck between there being enough time to return to clinical rotations after the course, to give participants the opportunity to reflect on the impact of the training and potential influence on their real-world practice, while still being in recent enough memory to be accurately recalled.

The majority of interviews (8 of 11) were face-to-face, 3 interviews were conducted by telephone at the request of the participants. All reasonable efforts were made to accommodate the participants busy working schedules and their preferences for location/interview modality. All interviews were conducted by HJ (orthopedic resident undertaking doctoral research) and recorded using a digital voice recorder. All interviews took place in a location familiar to the participants, and the interviewer was known to all participants as acquaintance and peer. Previous work has shown that matching the major social characteristics of the interviewer and interviewee is an important determining factor in the effectiveness of the interview.²⁰ Great care was taken to avoid imposing implicit interviewer biases on the participants, or steering the interview based on them, and to remain as objective and neutral as possible. The duality of the participant-interviewer peer relationship was managed mindfully and all participants were assured of confidentiality and gave their permission for the interview to be recorded and analyzed for research purposes. A prepiloted topic guide was used to structure the discussion (Appendix 1).

Data Analysis

An experiential thematic analysis approach was used for analysis, with a critical realist and postpositivist ontological and epistemological stance, respectively. This approach enabled complete focus on the participants own framing around issues, and their own terms of reference, allowing a fuller multifaceted understanding of the issues around the use of CST, in an exploratory and flexible manner that embraces the complexity of human experiences and perspectives. The analysis approach was structured around Braun & Clark’s checklist of criteria for “good thematic analysis.”²¹

The digital audio recordings were transcribed to a high level of orthographic detail and were rechecked against the original tapes to ensure accuracy. An initial process of reading and familiarization with the transcripts was followed by the start of the coding process and searching for themes. A complete coding strategy to identify “anything and everything” of interest within the entire dataset was used, generating a mixture of semantic and latent codes, in a recursive process over many weeks involving multiple revisions, until the entire dataset was completely coded. NVivo qualitative data analysis software version 11.4.3²² was used to collate relevant extracts for each theme. Once coding was felt to be complete, patterns were searched for within the coded data, from which to build themes. Themes generated during the analysis process were checked against each other and repeatedly referenced back to the original data set, to ensure they each had distinct scope and purpose, were faithful to the data, and that together they would provide a coherent and meaningful overview of key concepts in the data that addressed the research question.

RESULTS

Two key themes were evident: first, “factors driving learning from cadaveric simulation training,” which can be subdivided into intrinsic surgeon-driven factors, and external environmental factors (Table 1), and second, “added value of cadaveric simulation” (Table 2). This was the unifying concept of a cluster of findings relating to how cadaveric simulation can add value to training beyond that of other simulation modalities.

Intrinsic, Surgeon-Driven Factors

Self-perception of operative confidence influenced learning following CST. Those with low self-reported confidence in their operative skill appeared to make the greatest gains following the CST intervention

“for me before I started the course, bemiartbroplasty was my Everest and I think after that course I

TABLE 1. Summary of the Factors Driving Learning in Cadaveric Simulation

Intrinsic, Surgeon-Driven Factors	Extrinsic, Environmental Factors
Self-perception of skill level <ul style="list-style-type: none"> • Least confident participants appeared to make the most gains 	Paired learning <ul style="list-style-type: none"> • Practice assisting • Learn from colleague Multidisciplinary simulation <ul style="list-style-type: none"> • Experience being scrub nurse • Alternative perspectives Perform operations in their entirety <ul style="list-style-type: none"> • Fluency • Momentum Intensive consultant supervision <ul style="list-style-type: none"> • One-to-one • Real time feedback • Superspecialized • Tips/tricks • Unusual approaches Becoming through doing <ul style="list-style-type: none"> • Nontechnical skills • Highly realistic space, time and costuming • Sociocultural, “unspoken” lessons of surgery • Stirring to practice
Willingness to “buy-in” to the simulation <ul style="list-style-type: none"> • Suspension of disbelief • Staying in “character” 	
Pushing boundaries <ul style="list-style-type: none"> • Move out of comfort zone • Not ask for help as often • Safe space to make mistakes 	
Timing of delivery <ul style="list-style-type: none"> • Participant at correct stage of training for maximal benefit—early/mid-PGY4 • Do CST just before start of trauma rotation 	

was much more confident in approaching it” (Participant 12)

“I’m just much more happy and confident [having done the CST course] that I’ll be competent to do it [the operation]” (Participant 1)

The ability to “buy into” the simulation exercise and behave as if operating on a live patient was important for optimal learning and the ability to do this successfully varied between individuals. One participant had

TABLE 2. Summary of the “Added Value” of Cadaveric Simulation. Superiority of CST as compared other simulation modalities

Anatomical fidelity

- Soft-tissue envelope
- Neurovascular hazards
- Tissue tension
- Haptic feedback

“Dress rehearsal” for real surgery

- Whole learning, enabling direct skill transfer to the operating theater
- Can build on and refine foundational skills learnt in low-fidelity simulation
- Becoming through doing, “stirring into practice”

Offers solutions to some of the barriers to learning in the operating theater

- No time pressure
- No pressure from anesthetist/other theater staff
- No risk to patients

previously struggled with “suspension of disbelief” in the low-fidelity setting and appreciated the value of the high fidelity of the CST in achieving an immersive experience

“so in a [low-fidelity simulation setting] workshop, it is much easier to slip into not quite doing it properly, such as a soft tissue guide, because there isn’t soft tissue to worry about...suddenly the whole illusion breaks down” (participant 8)

An ability to push one’s own learning boundaries within the safety of a simulation exercise enabled learners to gain maximal benefit from the CST intervention. For example, whereas in real life, an inexperienced trainee often seeks reassurance from the trainer before progressing through each stage of the operation, in the simulation exercise, the trainee can move beyond their comfort zone under careful-guided supervision more confidently and improve their operative fluency without the risk of causing patient harm.

Within this sense of safety, and yet investment in the environmental fidelity of the simulation, participants reported valuing the opportunity to have the time to perform the operation in its entirety, without the usual time pressures of the operating theater

“so that [the CST intervention] was a great place to just do a procedure and not be cut there, cut there, cut there, just crack on, do it and it doesn’t matter if you get it wrong” (Participant 3)

“in an environment that wasn’t time pressured or, no external pressures, the anesthetist or something like that or an unwell patient. . .was very helpful indeed and from a learning point of view I think I learnt more in the cases I did there [the CST intervention] than the vast majority of cases that I do in theatre” (Participant 7)

“I got to the end and I thought ‘Ah I haven’t actually asked for help’ as supposed to in a real situation. . .I will be constantly looking for reassurance that the guide wire is in the right place, that you know, everything was, everything was right. . .and almost asking permission before going onto the next step” (Participant 6).

Correctly timing the delivery of the CST intervention within training was perceived to be a crucial intrinsic, learner-dependent factor to its success. Most of the participants felt that delivery of the course in its original format was best suited to the beginning or middle of PGY4. This was because at the start of PGY4 level, within the UK training system, the participants did not have much independent operative experience. Most participants reported encountering significant practical and logistical difficulties accessing the operating theater to receive conventional training, and yet competence in these index procedures was expected to be achieved to progress on to PGY5 (Specialist Registrar) level—thus, a paradox exists between the demands of the curriculum and the realities of the daily working environment these doctors find themselves in. This was a consistent finding across participants at PGY4 level who felt the course was appropriately timed for their stage of training, and participants at PGY5 level at the time of the intervention felt the course would have been of greater benefit to them a year earlier.

“I felt that the level I went into it at the beginning of PGY 4 was perfect because you, you’ve had a bit of time in trauma theatre, but maybe not as much independence as the senior trainees. . .and then getting a lot of confidence from having seen the four index procedures [on the CST course] (Participant 16)

“middle of PGY 4 [would be ideal], those procedures are, apart from the fasciotomies, I think they are all essentials for becoming a registrar [PGY 5]” (Participant 2)

Timing of delivery with respect to the commencement of a trauma rotation was also raised, with the course being more useful if it were delivered before the placement begins (as compared to 6 weeks in, as it was in this study),

and that there was a perceived risk of learned skill attrition if the course was delivered too early.

“If I had it just before [trauma rotation] that would be even better. . .you need to be doing it relatively fresh, within four to six weeks” (Participant 3)

Extrinsic, Environmental Factors

Participants were paired during the CST training intervention, and while one was operating as “first” surgeon, their partner was assisting or acting as scrub nurse if no assistant was required. The paired-learning nature of the CST intervention was perceived as valuable, as there was the opportunity to learn from the experience of a colleague partner

“I think having. . .two participants working together was very useful because you see one [procedure], your colleague doing it and then you do it yourself, you can kind of learn from each other and even if you do make any pitfalls you can kind of learn from that experience, and think what you would have done differently and so I think learning from each other is a really good thing” (Participant 15)

“It was nice to learn from your partner aswell, so if you saw a case that, for example, you weren’t particularly sure about then the next day you had opportunity to do it” (Participant 16)

The multidisciplinary environment of the CST intervention was perceived as valuable and enhanced learning, both in terms of enabling dialog about their performance between allied health professionals (scrub staff and radiographers) and through the opportunity to assume the role of scrub nurse. This gave participants an insight into the role the scrub nurses play in ensuring the smooth progress of an operation, and furthermore to improve their own knowledge of the sequence of steps in a procedure, as the scrub nurse is required to anticipate the next stage of a procedure and have the correct instruments to hand

“Actually what I found very helpful was being a scrub nurse and watching and anticipating and giving them the next thing and the next thing” (Participant 8)

“I liked the way that we had the theatre staff come in aswell, it was really good to get their opinions on things. . .I think that was very good” (Participant 10)

“. . .your interaction with your team, having that one on one feedback as well as you were doing the process [was beneficial to learning]” (Participant 16)

One of the key extrinsic features of CST that drive learning is the opportunity to perform operations in their entirety as first surgeon, and in an intensive way; that is, several successive operations in a short overall period of time. This would not typically be encountered in the real-life operating theater environment, because of the service demands placed upon trainee surgeons to process admissions, medically manage inpatients, and deal with administration. The participants valued this opportunity and felt that it enabled them to progress their skills more quickly than usual

“[CST] gives you an opportunity to do a lot of operations in a short period of time as first surgeon” (Participant 6)

“I think [the CST course] was the first time I had ever done a hemiarthroplasty entirely on my own” (Participant 12)

Another key feature driving learning was the intensive nature of the supervision during the CST intervention. Each operation was supervised by a consultant/attending, who provided real-time feedback and guidance, and were given the scope to challenge participants as they judged appropriate. Having intensive supervision also helped the learners maintain the fidelity of the simulation, as the faculty helped maintain the illusion that they were in the “real” operating theater through their nonverbal cues and behaviors

“it keeps you switched on and stops you lapsing, so [you aren’t] doing the ‘oh in real life’ I would have done this’, that’s really helpful” (Participant 8)

“to have high quality teachers one-on-one was fantastic... [the CST course] was an excellent way of learning, having the consultant [attending] stood over your shoulder which is something you might not have in the actual theatre itself” (Participant 3)

“[amongst] the things I found most helpful [about the CST] was you were getting one-on-one consultant level teaching” (Participant 7)

The US convention faculty were allocated stations according to their subspecialist interests, and their expertise was highly valued by the participants

“having an ankle surgeon [specialist] at the ankle station was good, because ankles can always be a bit fiddly and people have certain tips and tricks, it was very helpful” (Participant 15)

“I think it [the CST intervention] was really well thought out” (Participant 10)

There was recognition that simulating the complexities of the real-life operating theater was extremely difficult, and that CST was the best available simulation modality to try and achieve this replication

“its very hard to simulate training in orthopaedics... certainly cadaveric training is probably the only way you’re going to be able to do that” (Participant 2)

Some aspects of the CST intervention compromised the fidelity of the training experience, in particular, the specimens moved around during the hemiarthroplasty procedure (normally the patients’ body weight and positioning aids prevent this in real life). Waist-to-toe-tip specimens were used rather than whole cadavers, and so these obviously weighed less than a whole body. This also negatively affected the realism of relocating the hip once the implant was in position for similar reasons.

“for the hemiarthroplasty they [the cadavers] were just moving around a little bit and it wasn’t as realistic as you had in theatres” (Participant 16)

“having just the one leg for hemiarthroplasty made setting up quite difficult and when you tried to relocate the hip” (Participant 10)

The Added Value of Cadaveric Simulation

With cadaveric simulation being so much more expensive to deliver¹⁶ than other lower fidelity, inorganic types of simulation, it needs to bring additional benefits that these cheaper alternatives do not, to justify the cost. Participants were asked if there were features of the CST which were particularly useful to them in developing their skills as surgeons-in-training, and whether they felt CST offered value beyond that of other simulation modalities.

Anatomical Fidelity

Anatomical fidelity, the presence of a soft-tissue envelope and neurovascular structures as seen in life, were reported by participants as features peculiar to CST that could not be found elsewhere. Intuitively, deceased human bodies offer the most realistic representation of living anatomy, which is extremely difficult, if not impossible, to replicate by other means. Where visual representation of anatomy can be achieved by use of sophisticated computer and virtual reality programming, the haptic, tactile feedback and “tissue tension” experienced in cadaveric simulation is unparalleled.

“tissue tension is something that is quite unique to the cadavers really” (Participant 8)

“technical skills, in going through [dissecting] various layers, you can’t simulate [that] with dry bone” (Participant 7)

“life-like I suppose, with tactile feedback” (Participant 1)

“I think in terms of how high fidelity it was compared with what you normally do [in real life], it was very close” (Participant 16)

The presence of a soft-tissue envelope made the educational experience much more valuable for participants as they had to navigate the neurovascular hazards as they would in real life, and they could not “cheat” by obtaining direct visualization of the bone, as is possible in low-fidelity benchtop models such as sawbones. They were therefore more invested in the authenticity of the simulation experience, which became immersive, and led to other cognitive benefits.

Within the immersiveness of the experience, participants “bought into” the realism and began to behave as they would in the operating theater. This investment in the simulation exercise revealed another important area of the added value of CST—consequence and patient safety. Participants, while immersed in the realism of the training experience, knew that there were no real consequences to their making a mistake, and there was no risk to patient safety. In treating the exercise as a “full dress rehearsal” for real-life operating, the participants felt confident to push their own boundaries and progressed their learning as a result.

“it doesn’t matter if you get it wrong” (Participant 3)

“I think its good for the trainee, because they go through all the steps [in CST], they make sure they feel happy and confident, they’ve gone through the motions, and they can consolidate that on a cadaver first...and its good for patients because they get someone [a surgeon], they’re not practicing on a real person” (Participant 15).

Skills learnt in the CST intervention were directly transferred to the operating theater, as result of this “dress rehearsal” opportunity. There was no need for the participants to aggregate or embellish their learning before taking it to the operating theater, the learning from CST was whole, or complete.

For example, 1 participant described how a week after attending the CST course, they had been on-call over the weekend and been asked to perform a lower limb fasciotomy. The participant describes how there was no-

one available to supervise them performing the procedure, but having completed the procedure on the cadaver during the CST course the previous week, they felt confident of the surgical landmarks and hazards, in a way that reading about the techniques in a textbook would not have achieved.

“I knew where the perforators would be, it all went very smoothly, and its one of those things where if you’ve never done something before you can read it in a book, but you’re not going to be sure of yourself, and I think that once I’ve done something [on the cadavers] I know I can do it, then I’m just much more happy and confident that I’ll be competent to do it. . .doing that on a specimen rather than a person, in that situation [the fasciotomy] especially because its not something that you see every day, I might not do one for another few years, it’s very valuable” (Participant 1).

Another participant described how their supervising attending knew that they had recently successfully completed a fasciotomy procedure during the CST course, and so when a real case was encountered a few weeks later, they were happy to let the resident perform the entire operation on the patient, confident in the knowledge that the resident had previously achieved competence in the simulated environment.

“I definitely used what I learnt [on the course]. . .it’s definitely made a difference to training” (Participant 16)

Real-Time Feedback

Participants valued the real-time feedback and “tips and tricks” that supervising faculty provided during the CST course, and the opportunity to complete workplace-based assessments from their operative performances

“I really liked the one on one consultant [attending] feedback. I thought it was a really nice way of doing things” (Participant 16)

“it’s quite hard in day to day training sometimes to get the consultants to sit down and do the forms properly and give you constructive feedback. . .often they will say ‘oh just fill it in and send it to me’, but they had to do it properly [during the CST intervention]” (Participant 6)

“people have certain tips and tricks. . .it was very helpful” (Participant 15)

“for me, the best bit about cadaveric training is getting to do things [approaches] that you don’t normally do” (Participant 2)

Timing of Training Delivery

The timing of CST with respect to delivery of other low-fidelity simulation training opportunities was also explored with participants, with a particular emphasis toward understanding whether CST has an adjunctive or replacement role when compared with low-fidelity simulation.

As a prerequisite to completion of PGY4, all UK orthopedic residents must undertake the AO Foundation Basic Principles of Fracture Management Course.²³ This is an interactive course which teaches the basic concepts of stability, physiology of bone healing, and reduction and fixation techniques for simple fractures using low-fidelity, plastic bone simulation. Given the consensus among participants that the CST intervention was best delivered in the middle of the PGY4 year, that is, within the accepted timeframe for AO course completion, it was interesting to explore whether they felt that CST carried most educational benefit when delivered before or after the AO course.

Participants reported that the CST was most beneficial after they had grasped the basic principles of fracture management via the AO course, and that the more sophisticated simulation environment of CST allowed them to build on what they had already learnt in the low-fidelity environment. The AO course is very valuable as a first introduction to the principles and surgical instrumentation, which do not require the expense of CST to impart to surgeons-in-training.

“dry bones are really good, I think for basic principles and just getting familiar with technique and equipment. . .and then being able to apply those basic principles [in CST].if you like, a higher level of simulation” (Participant 7)

Participants were also asked about their opinions on whether there is a role for embedding CST within the curriculum, making it routinely accessible to surgeons-in-training as part of their formal teaching program. The response was very much in favor of this approach, that CST should be centrally funded and provided free of charge within residency programs, and was a tremendous yet presently underutilized training tool.

“Absolutely. Absolutely, I think it’s the way forward really” (Participant 15)

“It’d be a big loss if you weren’t able to build it into the curriculum” (Participant 16)

“they [the courses] should be delivered in region, for free to your trainees at the appropriate time” (Participant 3)

“I think having cadaveric [simulation] training is an unbelievable privilege for us and really, really useful” (Participant 10)

“I thought it [the CST intervention] was perfect, it was fantastic and I’m so lucky to be part of it” (Participant 12)

Value of CST in Developing Nontechnical Skills

In Cleland et al,¹⁹ a rapid ethnographic study of 2 surgical boot-camp training courses delivered to early stage-postgraduate surgical trainees was undertaken. The study aimed to understand the sociocultural influences of this intensive training and the wider implications for simulation-based education. The authors found that intensive boot-camp style training (of which CST is a variant) is “as much about social and cultural processes” as it is about “individual, cognitive and acquisitive learning.”¹⁹ These findings build on previous work examining how surgeons-in-training “become through doing.” Prentice, in an ethnography examining how medical students and junior doctors learn surgical skills in the operating theater, states that in order to gain a full understanding of how a “resident comes to embody the knowledge, skills and values of a surgeon requires understanding how social milieu and guided practice interact.”²⁴ Prentice describes the “guided physical training in the operating room” as embodying the “technical and social lessons of surgery,” even where the skills being taught are purely “technical.” Her findings reported that technical skill is only “20 percent” of the overall skillset required of a surgeon, with unspoken “tacit” knowledge, clinical judgment, and moral behavior forming a substantial part of what is required of a surgeon, beyond technical proficiency.²⁴

Previous qualitative work examining surgical practice has “shown surgical action in detail, but have little to say about how surgical trainees learn to fit themselves into the team, how they take on increasing levels of responsibility and how they develop the moral qualities of a surgeon.”²⁴ Prentice attempts to “unpack the unspoken lessons of surgery” by framing it within Bourdieu’s (Bourdieu 1977 cited by Prentice.²⁴) discussion of the “symbolically structured environment.” According to Bourdieu, the “structures of an environment build particular organizing principles, habits and the ways of being into the minds and bodies of cultural actors” (surgeons in this instance), and the symbolic structured environment (the operating theatre) exerts an anonymous, pervasive, pedagogic action.²⁴ Prentice elaborates on this with reference to surgical training, that in the operating theater the highly ritualized “space, time and costuming

provide structuring effects that make the imitation of a surgeon's actions and attitudes have meaning."²⁴ This helps create a positive economy in learners by instilling "the social hierarchy of the operating room," a hierarchy that "places surgeons in the centre, and gradually (with increasing experience) moves surgical residents into full participation at the centre of the action."

The value of CST, therefore, goes beyond provision of technical skills training. The ultrahigh environmental fidelity of CST replicates the "symbolic structured environment" of the operating theater and allows the learner to begin to become socialized in the practice of surgery and to "become through doing," learning both technical and nontechnical skills in a highly realistic environment, which itself exerts a "pervasive pedagogic action."

DISCUSSION AND IMPLICATIONS

Context of Findings Within Existing Evidence

Cleland's ethnographic study¹⁹ of a UK-based simulation "boot-camp" for PGY3 residents described 3 broad areas of educational gains following the training; technical (and nontechnical) skills, "cultural capital," which the authors describe as "resources in the form of learning what knowledge, skills, and values were needed to succeed in the surgical training system" and "social capital," in terms of extending their mentoring network.¹⁹ The authors acknowledge finding evidence of a distinction between the explicit and "hidden" curriculum within the boot-camp environment, with the latter adding value to the training by facilitating "enculturation and socialization into surgical training." They state that because this intensive training environment supports "both formal skills learning and informal learning about how to be a surgeon through social and cultural processes," it is important that simulation-based training program developers and researchers "address the social and cultural aspects of learning when planning similar enterprises," as educational interventions do not occur in "social, historical or cultural isolation."¹⁹

Their conclusion aligns with the findings of Jensen's²⁵ study of how medical students learn in the operating theater that the phenomena of surgical learning can be perceived as "instances of transformation in and among social practices," that students learn by "participating in the practice of providing high quality care," beyond simply technical skill acquisition in isolation, and the overall aim therefore is to teach "students to be surgeons instead of teaching them to perform surgery."²⁵

Our analysis shows that the sociocultural features of CST were valued by participants, it helped with preparing them for the PGY4-5 transition and developing their

professional identities and confidence as surgeons, and that the ultrahigh-fidelity nature of the simulation had additional, nuanced "cultural capital" benefits beyond the more obvious remit of developing technical skill acquisition. These benefits are particular to CST, as a consequence of successfully replicating the symbolically structured environment of the operating theater, with its associated "pervasive pedagogic action" in developing both technical and nontechnical "tacit" social skills and knowledge.²⁴

Strengths and Limitations

This study has several strengths; it is the first in-depth qualitative study (to our best knowledge) on the role of CST for training surgeons and thus adds to the existing evidence base in this area. We have followed Yardley's "open-ended, flexible" quality principles (Yardley 2000 cited by Braun & Clarke²⁶) in conducting this study, which "represent one of the most successful attempts to develop theoretically neutral validity criteria in qualitative research."²⁶ Commitment and rigor has been demonstrated by a thorough data collection and analysis procedure and by in-depth engagement with the research topic on both a professional and personal level. Transparency and coherence are also central to robust qualitative research practice, and we have endeavored to demonstrate this by presenting a clear analysis that is faithful to the dataset and theoretical framework, and reflexive in acknowledging the role that HJ has had, as both researcher and peer colleague of the participants, in shaping the research.

This study also has several weaknesses. The participants were all from one training region in the United Kingdom (West Midlands) and were individuals who had agreed to take part in the educational trial, and thus might represent a particularly motivated cohort of residents who are interested and engaged in simulation training research.

The interviews were conducted 6 months after the CST intervention and represent the participants' experiences in training at that point in time. Ideally, if resources had permitted, it would have been helpful to have repeated the interviews at a later stage, to further explore the longitudinal nature of the impact of CST, and to also interview the attending faculty involved in delivering the CST, to gain an understanding of their perspectives.

Three of the interviews were conducted over the telephone, which may have shaped the data obtained from these participants, as "virtual" interview methods can mean that information can be lost or misconstrued in the absence of nonverbal cues that occur during the interaction of a face-to-face interview.²⁰ The telephone interviews were shorter in length than the face-to-face interviews, and the data generated is likely to be different than that from a face-to-face encounter.

CONCLUSIONS

CST was highly valued by the resident participants in this study. Direct transfer of skills learnt in the cadaveric laboratory to the real operating theater was reported. CST can help offset the issues around accessing conventional training opportunities in the operating theater in the early stages of training, and it may serve to help residents achieve competency in index surgical procedures more quickly, with resultant patient safety benefits.

CST is advantageous over low-fidelity simulation and appears to offer “added value” for several reasons. The ultrahigh anatomical fidelity of cadavera presents an opportunity to practice operations with an unparalleled realism. The intensive consultant supervision with real-time performance feedback and opportunity to perform an operation in its entirety as first surgeon allows learners to push their own boundaries within the safety of a simulated environment. The multidisciplinary nature of CST allows the learner to experience the perspective of scrub nurse which enhances their knowledge of sequencing of operative steps, teaches anticipatory skills and team working.

CST has an adjunctive role alongside conventional surgical training and low-fidelity simulation. The middle of the PGY4 year was reported to be the best time to deliver this training course in the context of this group of participants, and the CST course added most value when delivered after low-fidelity simulation training, to build on the foundational skills already learnt.

CST has value beyond merely acquiring technical skills. Through the ultrahigh environmental fidelity of the simulation, the “pervasive pedagogic action” of the symbolically structured environment of the operating theater can be recreated, enabling surgeons-in-training to gain a myriad of nontechnical skills and become “stirred into practice” by learning the values, behaviors, and tacit knowledge required for surgical practice. CST may therefore be a promising candidate in the drive to reform surgical training within the current climate.

An important direction for future work in this area would be to further explore the role of CST in the acquisition of nontechnical skills. There is also a need to explore the experiences of the attending trainers in delivering CST.

STUDY DESCRIPTION

This was a qualitative research study using semistructured interviews with PGY3-6 orthopedic residents who had undertaken an intensive CST course to explore the impact the training had on their learning.

REFERENCES

1. Temple J. Time for Training. A Review of the Impact of the European Working Time Directive on the Quality of Training. London: Medical Education England; 2010.
2. Greenaway D. Securing the Future of Excellent Patient Care: Final Report of the Independent Review, London.
3. Curran I. A Framework for Technology Enhanced Learning. London: Department of Health; 2011.
4. Donaldson L. 150 Years of the Annual Report of the Chief Medical Officer. London: Department of Health; 2009.
5. A High Quality Workforce: NHS Next Stage Review. London: Department of Health; 2008.
6. Department of Health Council GM, *The Trainee Doctor*. 2011: London.
7. Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. *N Engl J Med*. 2006;355:2664–2669.
8. Hargreaves DH, Bowditch MG, Griffin DR. On-the-Job Training for Surgeons. London: The Royal Society of Medicine Press; 1997.
9. Walter AJ. Surgical education for the twenty-first century: beyond the apprentice model. *Obstet Gynecol Clin North Am*. 2006;33:233–236. vii.
10. Marron C, Byrnes C, Kirk S. An EWTD-compliant shift rota decreases training opportunities. *Bull R Coll Surg Engl*. 2005;87:246–248.
11. Pickersgill T. The European working time directive for doctors in training. *BMJ*. 2001;323:1266.
12. House J. Calling time on doctors' working hours. *Lancet*. 2009;373:2011–2012.
13. Shape of Training., Report from the UK Shape of Training Steering Group (UKSTSG). 2017: General Medical Council.
14. Royal College of Surgeons of England. Improving Surgical Training—Trainee Prospectus. London: The Royal College of Surgeons; 2018.
15. Mitha AP, Almekhlafi MA, Janjua MJ, Albuquerque FC, McDougall CG. Simulation and augmented reality in endovascular neurosurgery lessons from aviation. *Neurosurgery*. 2013;72(suppl_1):A107–A114.
16. Gilbody J, Prasthofer AW, Ho K, Costa ML. The use and effectiveness of cadaveric workshops in higher

surgical training: a systematic review. *Ann R Coll Surg Engl.* 2011;93:347–356.

17. Carey J, Minneti M, Leland HA, Demetriades D, Talving P. Perfused fresh cadavers: method for application to surgical simulation. *Am J Surg.* 2015;210:179–187.
18. Sugand K, Abrahams P, Khurana A. The anatomy of anatomy: a review for its modernisation. *Anat Sci Educ.* 2010;3:83–93.
19. Cleland J, Walker KG, Gale M, Nicol LG. Simulation-based education: understanding the socio-cultural complexity of a surgical training ‘boot camp’. *Med Educ.* 2016;50:829–841.
20. Braun V, Clark C. Successfully collecting qualitative data. Interactive data collection 1: interviews. *Successful Qualitative Research—A Practical Guide for Beginners.* SAGE Publications; 2013. p. 77–105.
21. Braun V, Clark C. Successfully analysing qualitative data. First analytic steps: familiarisation and data coding. *Successful Qualitative Research—A Practical Guide for Beginners.* SAGE Publications; 2013. p. 201–222.
22. NVivo Qualitative Data Analysis Software. QSR International Pty Ltd.; 2017 *Version 11.*
23. Available from: <https://www.aofoundation.org/Structure/Pages/default.aspx>.
24. Prentice R. Drilling surgeons: the social lessons of embodied surgical learning. *Sci Tech Human Val.* 2007;32:534–553.
25. Jensen RD, Seyer-Hansen M, Cristancho SM, Christensen MK. Being a surgeon or doing surgery? A qualitative study of learning in the operating room. *Med Educ.* 2018;52:861–876.
26. Braun V, Clark C. Quality criteria and techniques for qualitative research. *Successful Qualitative Research—A Practical Guide for Beginners.* SAGE Publications; 2013. p. 277–294.

APPENDIX 1

Interview Topic Guide for CAD:TRAUMA Participants

1. Demographic information about participant

Current post including specialty and hospital

- Stage of training

2. Topics for discussion (not exhaustive)

- Challenges of modern day surgical training
- Experiences of cadaveric simulation as an adjunct to training
- Factors that make good surgeons
- Preparedness for operating in real-life as a junior trainee surgeon

Questions will be framed to encompass four key domains – knowledge, opinion, feeling and experience (1)

Example knowledge questions:

- 1) Do you know what the European Working Time Directive says about the hours trainee surgeons are allowed to work?
- 2) Do you know how many index procedure* performances and work-based assessments you need to have logged by CCT**

Example opinion questions:

- 1) What do you think about modern surgical training in general? In Trauma & Orthopaedic surgery? In your hospital?
- 2) What do you think about using simulation to augment training? Cadaveric simulation in particular?
- 3) What do you think makes a good surgeon?

Example feeling questions:

- 1) What do you feel is the best way to train surgeons?
- 2) Do you feel that the CST training has benefited? How so?

Example experience questions:

- 1) Can you tell me about any times that you were not able to get access to the training opportunities you needed in the operating theatre? If yes -Why do you think this happened? If no – How is your training organized to avoid this happening?
- 2) When you have been doing operations in the recent past, have you felt well prepared?

* Index procedures are the key operations as defined in the curriculum

** Certificate of Completion of Training – the endpoint of surgical training

- (1) Lichtman, M. *Qualitative research in education: A user's guide.* 3rd Edition. London: Sage Publications; 2013

Systematic review of the current status of cadaveric simulation for surgical training

H. K. James^{1,2} , A. W. Chapman², G. T. R. Pattison², D. R. Griffin^{1,2} and J. D. Fisher¹

¹Clinical Trials Unit, Warwick Medical School, and ²Department of Trauma and Orthopaedic Surgery, University Hospitals Coventry and Warwickshire, Coventry, UK

Correspondence to: Dr H. K. James, Clinical Sciences Research Laboratories, University Hospital Coventry and Warwickshire, Clifford Bridge Road, Coventry CV2 2DX, UK (e-mail: )

Background: There is growing interest in and provision of cadaveric simulation courses for surgical trainees. This is being driven by the need to modernize and improve the efficiency of surgical training within the current challenging training climate. The objective of this systematic review is to describe and evaluate the evidence for cadaveric simulation in postgraduate surgical training.

Methods: A PRISMA-compliant systematic literature review of studies that prospectively evaluated a cadaveric simulation training intervention for surgical trainees was undertaken. All relevant databases and trial registries were searched to January 2019. Methodological rigour was assessed using the widely validated Medical Education Research Quality Index (MERSQI) tool.

Results: A total of 51 studies were included, involving 2002 surgical trainees across 69 cadaveric training interventions. Of these, 22 assessed the impact of the cadaveric training intervention using only subjective measures, five measured impact by change in learner knowledge, and 23 used objective tools to assess change in learner behaviour after training. Only one study assessed patient outcome and demonstrated transfer of skill from the simulated environment to the workplace. Of the included studies, 67 per cent had weak methodology (MERSQI score less than 10.7).

Conclusion: There is an abundance of relatively low-quality evidence showing that cadaveric simulation induces short-term skill acquisition as measured by objective means. There is currently a lack of evidence of skill retention, and of transfer of skills following training into the live operating theatre.

Paper accepted 4 July 2019

Published online 1 October 2019 in Wiley Online Library (www.bjs.co.uk). DOI: 10.1002/bjs.11325

Introduction

There is growing interest in the use of cadaveric simulation in postgraduate surgical training¹. The move to incorporate simulation into surgical training is driven by a need to improve training efficiency in the current climate of reduced working hours^{2,3}, financial constraint and emphasis on patient safety⁴. Cadaveric simulation is of particular interest, as it provides ultra-high-fidelity representation of surgical anatomy as encountered *in vivo*^{1,5}, authentic tissue handling⁶ and complex three-dimensional neurovascular relationships, which are difficult to appreciate in textbooks and almost impossible to replicate in synthetic models⁷. Cadaveric simulation offers the opportunity to practise an operation in its entirety with high environmental, equipment and psychological fidelity, thereby enabling the rapid acquisition of procedural skills⁵ and attainment of competence in a setting remote from patient

care⁸. With the current increase in availability of cadaveric training courses for surgical trainees⁹, a systematic evaluation of the evidence for their use is both timely and necessary.

The purpose of this review was to describe and evaluate the evidence for the use of cadaveric simulation in postgraduate surgical training.

Methods

This systematic review was conducted in accordance with the PRISMA guidelines^{10,11}; the review protocol was registered with PROSPERO (an international prospective register of systematic reviews)¹².

Search strategy and data sources

A literature search was conducted in January 2019 using MEDLINE (Ovid) (1946 to the present), CINAHL

(EBSCO) (Cumulative Index of Nursing and Allied Health Literature), Centre for Reviews and Dissemination Database, ISRCTN Registry, Cochrane Central Register of Controlled Trials, NHS Evidence, PubMed (1950 to the present), Embase (Ovid) (1947 to the present), Scopus, Australian Clinical Trials Registry and Google Scholar. Medical Subject Headings (MeSH) terms and text words from the MEDLINE search strategy (*Table S1*, supporting information) were adapted for other databases according to the required syntax.

Search results were limited to human subjects and the English language. Duplicates were removed, and retrieved titles and abstracts were screened for initial eligibility. Reference lists of included studies and old reviews were hand-searched to ensure literature saturation.

Selection criteria and data extraction

The initial eligibility screening criteria were: study participants were postgraduate doctors in training; there was exposure to human cadaveric simulation training; and there was an attempt at measuring the educational impact.

Studies were excluded at screening if they used animal cadaveric models, involved veterinary trainees, or were purely descriptive feasibility studies describing a cadaveric technique, with no assessment of the educational impact.

Abstracts that passed eligibility screening were retrieved in full text. Reference lists of full-text articles were examined for relevant studies, and those found by hand-searching were subject to the same eligibility screening process.

The data were extracted from the full-text articles using piloted data extraction forms, by two reviewers working independently. Data items collected included: participant characteristics (number, stage of training and specialty); study characteristics (single-centre *versus* multicentre, eligibility criteria defined, loss to follow-up); cadaveric training (intervention, cadaveric model used, skills taught, comparator group (where applicable)); assessment of educational impact (primary outcome measure, evidence of instrument validation, results summary (objective and subjective), post-test assessment and evidence of skill transfer (if applicable)).

Data analysis, quality assessment and evidence synthesis

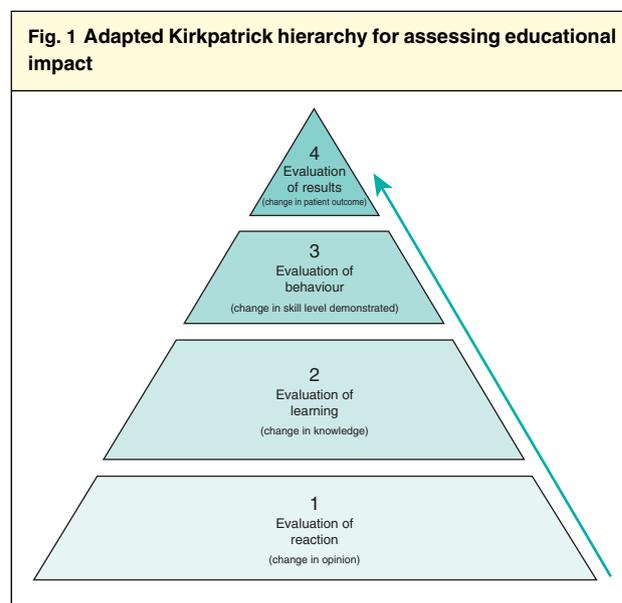
Included studies were assigned a level of evidence score using a modified version of the Oxford Centre for Evidence-Based Medicine (OCEBM) classification, which has been adapted by the European Association of

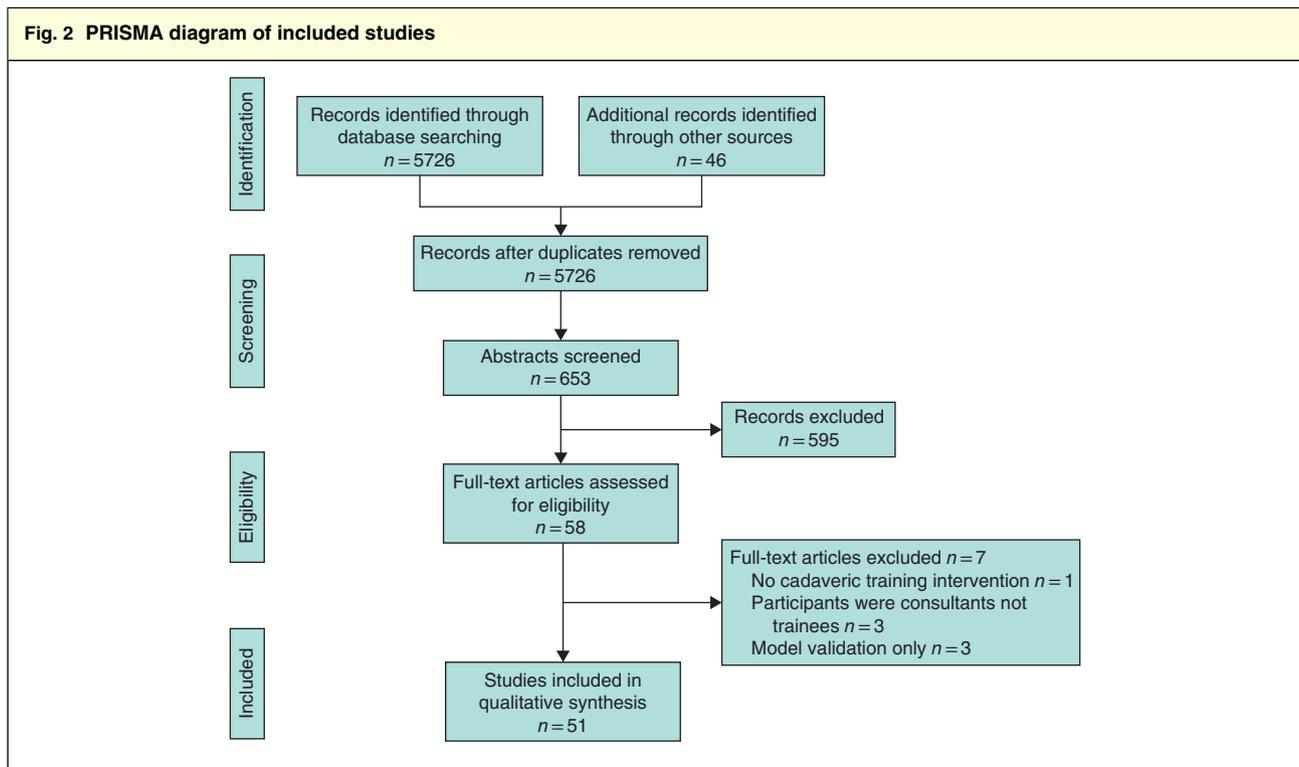
Endoscopic Surgery and is used widely in educational systematic reviews^{13,14}. Methodological rigour of included studies was scored using the Medical Education Research Quality Instrument (MERSQI), which is a previously validated assessment tool^{15–17} for quantitative appraisal of medical education research across six domains: study design, sampling, type of data, instrument validity, data analysis and outcome. The maximum score is 18 points. The mean MERSQI score of both independent assessors for each included study is reported.

A qualitative, narrative synthesis of evidence was undertaken, structured around an adapted Kirkpatrick's hierarchy for assessing the educational impact of a teaching intervention (*Fig. 1*).

Results

The initial search generated 5726 results, of which 5073 were clearly ineligible and rejected at title review (*Fig. 2*). A total of 653 abstracts were screened, 595 of which did not pass eligibility screening and were excluded. Some 58 articles were accessed in full text and reviewed carefully; one study was rejected at this stage as there was no cadaveric simulation training intervention, three were rejected as the study participants were consultants not trainees, and three studies were rejected as there was cadaveric model validation only with no assessment of educational impact. Fifty-one studies were included in the review, of which 47 were full-text original research articles and four were conference posters. The main characteristics of studies,





including OCEBM and mean MERSQI scores are shown in *Tables S2–S5* (supporting information)^{18–68}.

Study design and setting

Eight studies were RCTs, six were comparative cohort studies, and 37 were non-comparative cohort studies.

The majority of studies were from the USA (35 studies) and the UK (8), with the remainder from Canada (4), Australia (2) and one each from Germany and Japan. All studies, except one³³ were delivered at a single centre.

Participants

The number of participants in the included studies ranged from three to 390, totalling 2002 individual participants across 69 cadaveric training interventions, representing the breadth of surgical training grades.

Surgical speciality

In total, 12 surgical specialties and subspecialties were included (*Tables S2–S5*). Most studies were within general surgery (14), trauma and orthopaedic surgery (9) and neurosurgery (7). All studies were single-speciality.

Study quality

The mean MERSQI score was 9.4 (range 5–14). In terms of level of evidence, only two of 51 studies were OCEBM level 1b (RCT of good quality and adequate sample size with a power calculation), six studies were OCEBM 2a (RCT of reasonable quality and/or of inadequate sample size), six were OCEBM 2b (parallel cohort study), and 37 were OCEBM level 3 (non-randomized, non-comparative trials, descriptive research).

A linear relationship was observed between Kirkpatrick level and mean MERSQI score, suggesting that quality of evidence is linked with robust methodology.

Measurement of educational impact

An assessment of educational impact of the training intervention was made using objective measures in 28 of the 51 included studies, and using subjective measures only in the other 23 studies. Sixteen of the 28 studies that used objective outcome measures attempted to measure skill transfer after training.

Level 1: Reaction

Twenty-two of the 51 studies measured the educational impact of a cadaveric training intervention using subjective

measures of learner reaction/opinion. One¹⁸ of these studies used a comparative cohort design, comparing cadaveric-trained with virtual reality-trained participants (OCEBM level 2b), and 21^{19–39} were descriptive research studies using non-randomized, non-comparative methods (OCEBM level 3). Two^{19,29} of the Kirkpatrick level 1 studies attempted to measure skill transfer following the cadaveric training intervention. All 22 studies used participant questionnaires to assess learner reaction, most of which were purpose-designed³⁸ and not validated formally. The outcome measures included learner reaction with respect to simulation fidelity, learner opinion on the usefulness of the training, and change in operative confidence and self-perceived competency after the training. All level 1 studies reported a positive effect of the cadaveric simulation training as measured by learner reaction/opinion.

Level 2: Learning

Five studies assessed the educational impact of the cadaveric training intervention by measuring change in learner knowledge. One⁴⁰ of these studies was an RCT comparing cadaveric simulation training with a low-fidelity bench-top simulator, one⁴¹ was a cohort study comparing learning in cadaveric-trained participants with those who received didactic teaching materials only, and three^{42–44} were non-comparative cohort studies. Three studies^{40,43,44} used procedural knowledge scores as the primary outcome measure, and two^{41,42} used *viva voce* and oral checklist examinations.

Cadaveric simulation training made no difference to the postintervention test scores in the study of AlJamal and colleagues⁴⁰, but a significant improvement in overall examination scores was seen in the cadaveric-trained group by Sharma and co-workers⁴¹. Significant improvement in post-test knowledge scores was reported in the three non-comparative studies^{42–44} following cadaveric training.

Level 3: Behaviour

Twenty-three studies assessed the educational impact of a cadaveric training intervention by attempting to measure a change in learner behaviour. Objective assessment methods of learner behaviour were highly variable, and included operational metrics (such as procedure time, error rate, hand motion analysis, path length) and final product analysis. Various score-based methods were also used, including procedure scores, global rating scale (GRS), OSATS (Objective Structured Assessment of Technical Skills in Surgery) and the GOALS (Global Operative Assessment of Laparoscopic Skills) scale. Seven^{45–51} of the 23 studies

were RCTs and 16^{52–67} were cohort studies. Of the seven RCTs, three^{45–47} compared cadaveric simulation with no simulation training, and four^{48–51} compared cadaveric simulation with low-fidelity simulation.

Compared with no simulation training, cadaveric simulation-trained learners showed significant improvement in most of the tested skill domains^{45–47}. When comparing behaviour change after training in low-fidelity simulation-trained and cadaveric simulation-trained learners, the results were mixed. Camp *et al.*⁵¹ reported that cadaveric training was superior to virtual reality (VR) when teaching knee arthroscopy, with greater improvement in procedural rating scores and reduced procedure time seen in the cadaveric-trained compared with the VR-trained group. Sidhu and colleagues⁴⁹ reported that cadaveric training was superior to a bench-top simulator for teaching graft-to-arterial anastomosis, as measured by a task-based checklist (TBC), GRS and final product analysis (FPA). A greater benefit of the cadaveric training was seen in the more junior study participants.

Conversely, Anastakis and co-workers⁴⁸ compared behaviour change in cadaveric-trained, low-fidelity bench model-trained and written materials only-trained groups of learners performing basic general surgical skills, measured by procedural checklist scores and GRS. They found that the bench- and cadaveric-trained groups performed better than the written materials only group, and that performances of the cadaveric- and bench-trained groups were equivalent. Gottschalk *et al.*⁵⁰ compared the performance of cadaveric-trained, low-fidelity bench model-trained and 'no training' groups at cervical lateral mass screw placement using FPA. They found that, although both the cadaveric- and bench-trained groups outperformed the no training group, the bench-trained group had greater improvement in performance.

Of the 16 cohort studies measuring change in learner behaviour, five were comparative in design. Three studies^{53–55} compared inexperienced *versus* experienced performance, one⁵² compared behaviour change in cadaveric simulation-trained *versus* low-fidelity simulation-trained cohorts, and one⁵⁶ compared within-subject performance change after cadaveric simulation training. Eleven^{57–67} were non-comparative descriptive studies.

The primary objective of the three studies comparing inexperienced and experienced performance was construct validation of the simulator and/or assessment tools used in the studies. Zirkle and colleagues⁵³ found that, when performing cortical mastoidectomy in a cadaveric simulation setting, FPA did not correlate with trainee experience, but GRS and TBC scores did. Mednick and co-workers⁵⁵ also found that, when performing corneal rust ring removal,

FPA did not correlate with trainee experience, although procedure time did. Mackenzie *et al.*⁵⁴ compared preintervention, immediately after (less than 4 weeks) and delayed (12–18 months) intervention scores for cadaveric-trained, inexperienced learners with experienced ‘expert’ performance when undertaking lower-extremity vascular exposure, repair and fasciotomy. The outcome measures were TBC, GRS, error frequency and procedure time. The results showed that experienced performance was significantly better at all time points, that performance amongst the inexperienced group was highly variable, and that evidence of skill retention was seen at 18 months postintervention.

When comparing cadaveric simulation-trained and low-fidelity (VR) simulation-trained cohorts performing laparoscopic sigmoid colectomy, LeBlanc *et al.*⁵² reported that technical skills scores were better in the low-fidelity group.

Of the 11 non-comparative descriptive studies, all reported improvement in trainee performance after cadaveric simulation training, using a variety of outcome measures such as FPA, GRS and operational metrics.

Level 4: Objective measurement of educational impact by change in patient outcome

Only one⁶⁸ of the 51 studies included in this review assessed the impact of the cadaveric training intervention on real-world patient outcomes. Martin and colleagues⁶⁸ measured the impact of cadaveric training on the real-world performance of ‘core’ invasive skills (endotracheal tube insertion, chest tube insertion and venous cut-down) by eight surgical trainees during the first 3 months of a trauma rotation. The complication rate for all skills decreased significantly immediately and at 3 weeks after instruction ($P < 0.02$). Initial trauma resuscitation time after training decreased from approximately 25 to 10 min in 80 patients treated by the participants⁶⁸. The authors concluded that trainees’ skills improve rapidly with competency-based instruction (CBI), skills learnt through CBI in the laboratory can be translated to and sustained in the clinical setting, and CBI yields competent trainees who perform skills rapidly and with minimal complications.

Cadaveric models used for simulation training

A wide variety of cadaveric models were used in the included studies. Three studies^{26,35,67} used innovative techniques to perfuse or reconstitute cadaveric material for training purposes, to improve the fidelity of the simulation. These studies were all in the field of neurosurgery, and involved cannulation of the great vessels of

the neck to allow pulsatile perfusion of cadaveric heads with an artificial blood substitute. All reported very high learner satisfaction with the models, and recognition of the opportunity that live reconstitution offers for overcoming the criticism^{35,67} of conventional, non-perfused cadaveric material, in that it does not bleed and thus the simulation fidelity is limited for teaching procedures where bleeding is a potential major consequence.

Fresh cadavers were used in 12^{20,23,28,33,37,43,48,52,54,55,65,68} of the reviewed studies. These offer the most authentic tissue-handling fidelity⁶⁹, but have the significant disadvantage of rapid deterioration, and therefore a short time-window for their potential use. Use of fresh cadavers for simulation training relies on a regular, local system of body donation bequests, as they are typically used within 48 h of the donor’s death, and certainly no more than 7 days later, which places logistical and infrastructure challenges on training providers.

Fresh-frozen cadavers, used in 12^{18,21,27,30–32,39,42,45,46,56,57} of the studies, have gained popularity due to their versatility. The cadavers are non-exsanguinated, washed with antiseptic soap, and frozen to -20°C within 1 week of procurement⁷⁰. Typically, around 3 days before use they are gradually thawed at room temperature, retaining the realistic tissue-handling characteristics that are important for high-fidelity simulation. Fresh-frozen cadavers have the great advantage of being able to be refrozen and thawed at a later date, permitting reuse across multiple training interventions and thus maximizing potential use and cost-efficiency⁷⁰.

Soft-fix Thiel embalming techniques were used in two studies^{34,58}. This technique seeks a method of cadaveric preservation that preserves tissue-handling, enables longevity of specimen use, and avoids the occupational and environmental health risks associated with exposure to formaldehyde⁷¹. Organs and tissues retain their flexibility, and the colour of the tissue remains similar to that seen *in vivo*.

Only one study⁴¹ used traditional formalin-fixed cadavers. Formalin has the advantages of being relatively inexpensive and widely available, with a long history of use in preserving cadavers for the purposes of teaching anatomy⁷¹. It does, however, lead to changes in the colour, strength and tissue-handling characteristics of the cadaveric material⁷², which may limit its usefulness in surgical training. The study by Sharma *et al.*⁴¹ did not evaluate the fidelity of the simulation or discuss the rationale or impact on the educational value of the training as a result of using formalin-fixed material. Twenty studies provided no information on the type of cadaveric material used in the training intervention.

Almost half of the included studies (22 of 51, 43 per cent) assessed the impact of the cadaveric training using subjective measures only, representing the lowest level of impact in educational research (Kirkpatrick level 1). The second most prevalent category was studies measuring the impact of the training intervention by learner behaviour change (23 of 51, 45 per cent) (Kirkpatrick level 3). Of these, seven were RCTs and 16 were cohort studies. Of the 23 studies assessing behaviour change following cadaveric training, only one⁶⁵ measured behaviour change in the workplace; the rest measured behaviour change in the simulation laboratory. Only one⁶⁸ of the 51 studies actually measured a change in patient outcome as a result of the cadaveric training intervention; this is the highest level of impact assessment in educational research (Kirkpatrick level 4).

Discussion

The objective of this systematic review was to describe and evaluate evidence for the current use of cadaveric simulation in postgraduate surgical training. Fifty-one studies involving 2002 surgical trainees across 69 cadaveric training interventions were included. Although there was research activity encompassing the breadth of surgical specialties, most studies were within general surgery (14), trauma and orthopaedic surgery (9) and neurosurgery (7). The majority were conducted in the USA (35) and UK (8). A wide range of methodology was used. Eight of 51 studies were RCTs (OCEBM level 1b and 2a), six were parallel cohort studies (OCEBM level 2b) and 37 were non-comparative descriptive studies (OCEBM level 3).

There is evidence from three RCTs^{45–47} that cadaveric simulation training is superior to no training, yet the important question remains whether it is superior to low-fidelity simulation training. This is of interest because cadaveric simulation training is expensive, and therefore needs to have demonstrable superiority over less expensive alternatives. The four RCTs^{48–51} in this area revealed a mixed picture: two^{49,51} showed superiority of cadaveric simulation training, one⁴⁸ showed equivalence with low-fidelity bench-model training, and one⁵⁰ showed that cadaveric simulation training was inferior to the bench-model alternative.

The mean overall MERSQI score correlated well with the Kirkpatrick level: the higher the level of impact measured, the better the study methodology. Previous predictive studies¹⁷ have shown that studies with a MERSQI score of 10.7 or above are indicative of a methodologically strong study, likely to be accepted for publication. Some 67 per cent of the studies in this review had a MERSQI score

below 10.7, and therefore the majority of the included studies could be considered to have weak methodology. Methodological problems noted amongst the included studies were: predominance of single-site studies, lack of randomization, lack of comparator group, small sample sizes and underpowering, inadequate or absent reporting of descriptive statistics, overuse of a single group, and predominance of pretest/post-test assessment strategies to determine the impact of the training intervention, which can overestimate the observed effect size⁷³.

Several studies reported mixed-modality training interventions, making assessment of the cadaveric component of the training in isolation impossible. There was also inadequate or absent description of the cadaveric model used in more than one-third of the studies, which renders results pertaining to the face and content validity of training impossible to assess.

Half of the studies were published in the last 3 years, reflecting the recent explosion in popularity of cadaveric simulation training. Despite the clear attraction of cadaveric simulation training as measured by subjective means, there remains a dearth of evidence that there is retention and translation of skills learnt in the cadaveric laboratory into the operating theatre. The major ongoing challenge within educational research is demonstrating effective, sustained changes in learner behaviour, and improved patient outcome following a training intervention⁷³. These challenges are particularly acute in the field of cadaveric simulation because of the cost and infrastructure demands, and there is presently little evidence that surgical educators are rising to meet this challenge.

This review has shown that there is an abundance of relatively low-quality evidence indicating that cadaveric simulation may induce short-term skill acquisition as measured by objective means. Adequately powered studies are needed to show whether skills are retained and transferable into the operating theatre before major investment in cadaveric simulation for surgical training can be recommended.

Disclosure

The authors declare no conflict of interest.

References

- 1 Gilbody J, Prasthofer AW, Ho K, Costa ML. The use and effectiveness of cadaveric workshops in higher surgical training: a systematic review. *Ann R Coll Surg Engl* 2011; **93**: 347–352.
- 2 Pickersgill T. The European working time directive for doctors in training. *BMJ* 2001; **323**: 1266.

- 3 House J. Calling time on doctors' working hours. *Lancet* 2009; **373**: 2011–2012.
- 4 UK Shape of Training Steering Group (UKSTSG). *Shape of Training: Report from the UK Shape of Training Steering Group (UKSTSG)*; 2017. https://www.shapeoftraining.co.uk/static/documents/content/Shape_of_Training_Final_SCT0417353814.pdf [accessed 1 December 2018].
- 5 Reznick RK, MacRae H. Teaching surgical skills – changes in the wind. *N Engl J Med* 2006; **355**: 2664–2669.
- 6 Carey JN, Minneti M, Leland HA, Demetriades D, Talving P. Perfused fresh cadavers: method for application to surgical simulation. *Am J Surg* 2015; **210**: 179–187.
- 7 Sugand K, Abrahams P, Khurana A. The anatomy of anatomy: a review for its modernisation. *Anat Sci Educ* 2010; **3**: 83–93.
- 8 Curran I. *A Framework for Technology Enhanced Learning*. Department of Health: London, 2011.
- 9 Hughes JL, Katsogridakis E, Junaid-Siddiqi A, Pollard JS, Bedford JD, Coe PO *et al.* The development of cadaveric simulation for core surgical trainees. *Bull Roy Coll Surg Engl* 2018; **101**: 38–43.
- 10 Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M *et al.*; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015; **350**: g7647.
- 11 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Int Med* 2009; **151**: W-65–W-94.
- 12 National Institute for Health Research. *PROSPERO: International Prospective Register of Systematic Reviews*. <https://www.crd.york.ac.uk/prospero/> [accessed 14 December 2018].
- 13 Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB *et al.* Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc* 2005; **19**: 1523–1532.
- 14 Morgan M, Aydin A, Salih A, Robati S, Ahmed K. Current status of simulation-based training tools in orthopedic surgery: a systematic review. *J Surg Educ* 2017; **74**: 698–716.
- 15 Cook DA, Reed DA. Appraising the quality of medical education research methods: the medical education research study quality instrument and the Newcastle–Ottawa Scale – Education. *Acad Med* 2015; **90**: 1067–1076.
- 16 Smith RP, Learman LA. A plea for MERSQI: the medical education research study quality instrument. *Obstet Gynecol* 2017; **130**: 686–690.
- 17 Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's medical education special issue. *J Gen Intern Med* 2008; **23**: 903–907.
- 18 Sharma M, Horgan A. Comparison of fresh-frozen cadaver and high-fidelity virtual reality simulator as methods of laparoscopic training. *World J Surg* 2012; **36**: 1732–1737.
- 19 Dunnington GL. A model for teaching sentinel lymph node mapping and excision and axillary lymph node dissection. *J Am Coll Surg* 2003; **197**: 119–121.
- 20 Gunst M, O'Keeffe T, Hollett L, Hamill M, Gentilello LM, Frankel H *et al.* Trauma operative skills in the era of nonoperative management: the trauma exposure course (TEC). *J Trauma* 2009; **67**: 1091–1096.
- 21 Reed AB, Crafton C, Giglia JS, Hutto JD. Back to basics: use of fresh cadavers in vascular surgery training. *Surgery* 2009; **146**: 757–763.
- 22 Lewis CE, Peacock WJ, Tillou A, Hines OJ, Hiatt JR. A novel cadaver-based educational program in general surgery training. *J Surg Educ* 2013; **69**: 693–698.
- 23 Shekter CC, Kane JT, Minneti M, Garner W, Sullivan M, Talving P *et al.* Incorporation of fresh tissue surgical simulation into plastic surgery education: maximizing extraclinical surgical experience. *J Surg Educ* 2013; **70**: 466–474.
- 24 Gasco J, Holbrook TJ, Patel A, Smith A, Paulson D, Muns A *et al.* Neurosurgery simulation in residency training: feasibility, cost, and educational benefit. *Neurosurgery* 2013; **73**: 39–45.
- 25 Jansen S, Cowie M, Linehan J, Hamdorf JM. Fresh frozen cadaver workshops for advanced vascular surgical training. *ANZ J Surg* 2014; **84**: 877–880.
- 26 Pham M, Kale A, Marquez Y, Winer J, Lee B, Harris B *et al.* A perfusion-based human cadaveric model for management of carotid artery injury during endoscopic endonasal skull base surgery. *J Neurol Surg B Skull Base* 2014; **75**: 309–313.
- 27 Ahmed K, Aydin A, Dasgupta P, Khan MS, McCabe JE. A novel cadaveric simulation program in urology. *J Surg Educ* 2015; **72**: 556–565.
- 28 Aydin A, Ahmed K, Khan MS, Dasgupta P, McCabe J. The role of human cadaveric procedural simulation in urology training. *J Urol* 2015; **193**: e273.
- 29 Liu JK, Kshetry VR, Recinos PF, Kamian K, Schlenk RP, Benz EC. Establishing a surgical skills laboratory and dissection curriculum for neurosurgical residency training. *J Neurosurg* 2016; **123**: 1331–1338.
- 30 Clark MA, Yoo SS, Kundu RV. Nail surgery techniques: a single center survey study on the effect of a cadaveric hand practicum in dermatology resident education. *Dermatol Surg* 2016; **42**: 696–698.
- 31 Hanu-Cernat L, Thanabalan N, Alali S, Viridi M. Effectiveness of cadaveric surgical simulation as perceived by participating OMFS trainees. *Br J Oral Maxillofac Surg* 2016; **54**: e126.
- 32 Calio BP, Kepler CK, Koerner JD, Rihn JA, Millhouse P, Radcliff KE. Outcome of a resident spine surgical skills training program. *Clin Spine Surg* 2017; **30**: E1126–E1129.
- 33 Lorenz R, Stechemesser B, Reinhold W, Fortelny R, Mayer F, Schroder W *et al.* Development of a standardized curriculum concept for continuing training in hernia surgery: German Hernia School. *Hernia* 2017; **21**: 153–162.

- 34 Nobuoka D, Yagi T, Kondo Y, Umeda Y, Yoshida R, Kuise T *et al.* The utility of cadaver-based surgical training in hepatobiliary and pancreatic surgery. *J Hepatobiliary Pancreat Sci* 2017; **24**: A338.
- 35 Pacca P, Jhawar SS, Seclen DV, Wang E, Snyderman C, Gardner PA *et al.* 'Live cadaver' model for internal carotid artery injury simulation in endoscopic endonasal skull base surgery. *Oper Neurosurg (Hagerstown)* 2017; **13**: 732–738.
- 36 Takayesu JK, Peak D, Stearns D. Cadaver-based training is superior to simulation training for cricothyrotomy and tube thoracostomy. *Intern Emerg Med* 2017; **12**: 99–102.
- 37 Weber EL, Leland HA, Azadgoli B, Minneti M, Carey JN. Preoperative surgical rehearsal using cadaveric fresh tissue surgical simulation increases resident operative confidence. *Ann Transl Med* 2017; **5**: 302.
- 38 Bertolo R, Garisto J, Dagenais J, Sagalovich D, Kaouk JH. Single session of robotic human cadaver training: the immediate impact on urology residents in a teaching hospital. *J Laparoendosc Adv Surg Tech A* 2018; **28**: 1157–1162.
- 39 Chouari TAM, Lindsay K, Bradshaw E, Parson S, Watson L, Ahmed J *et al.* An enhanced fresh cadaveric model for reconstructive microsurgery training. *Eur J Plast Surg* 2018; **41**: 439–446.
- 40 AlJamal Y, Buckarma E, Ruparel R, Allen S, Farley D. Cadaveric dissection *vs* homemade model: what is the best way to teach endoscopic totally extraperitoneal inguinal hernia repair? *J Surg Educ* 2018; **75**: 787–791.
- 41 Sharma G, Aycart MA, Najjar PA, van Houten T, Smink DS, Askari R *et al.* A cadaveric procedural anatomy course enhances operative competence. *J Surg Res* 2016; **201**: 22–28.
- 42 Mitchell EL, Sevdalis N, Arora S, Azarbal AF, Liem TK, Landry GJ *et al.* A fresh cadaver laboratory to conceptualize troublesome anatomic relationships in vascular surgery. *J Vasc Surg* 2012; **55**: 1187–1194.
- 43 Robinson WP, Doucet DR, Simons JP, Wyman A, Aiello FA, Arous E *et al.* An intensive vascular surgical skills and simulation course for vascular trainees improves procedural knowledge and self-rated procedural competence. *J Vasc Surg* 2017; **65**: 907–915.e3.
- 44 Hazan E, Torbeck R, Connolly D, Wang JV, Griffin T, Keller M *et al.* Cadaveric simulation for improving surgical training in dermatology. *Dermatol Online J* 2018; **24**: 15.
- 45 Sharma M, Macafee D, Horgan AF. Basic laparoscopic skills training using fresh frozen cadaver: a randomized controlled trial. *Am J Surg* 2013; **206**: 23–31.
- 46 Sundar SJ, Healy AT, Kshetry VR, Mroz TE, Schlenk R, Benzel EC. A pilot study of the utility of a laboratory-based spinal fixation training program for neurosurgical residents. *J Neurosurg Spine* 2016; **24**: 850–856.
- 47 Chong W, Downing K, Leegant A, Banks E, Fridman D, Downie S. Resident knowledge, surgical skill, and confidence in transobturator vaginal tape placement: the value of a cadaver laboratory. *Female Pelvic Med Reconstr Surg* 2017; **23**: 392–400.
- 48 Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M *et al.* Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg* 1999; **177**: 167–170.
- 49 Sidhu RS, Park J, Brydges R, MacRae HM, Dubrowski A. Laboratory-based vascular anastomosis training: a randomized controlled trial evaluating the effects of bench model fidelity and level of training on skill acquisition. *J Vasc Surg* 1999; **45**: 343–349.
- 50 Gottschalk MB, Yoon ST, Park DK, Rhee JM, Mitchell PM. Surgical training using three-dimensional simulation in placement of cervical lateral mass screws: a blinded randomized control trial. *Spine J* 2015; **15**: 168–175.
- 51 Camp CL, Krych AJ, Stuart MJ, Regnier TD, Mills KM, Turner NS. Improving resident performance in knee arthroscopy: a prospective value assessment of simulators and cadaveric skills laboratories. *J Bone Joint Surg Am* 2016; **98**: 220–225.
- 52 LeBlanc F, Champagne BJ, Augestad KM, Neary PC, Senagore AJ, Ellis CN *et al.* A comparison of human cadaver and augmented reality simulator models for straight laparoscopic colorectal skills acquisition training. *J Am Coll Surg* 2010; **211**: 250–255.
- 53 Zirkle M, Taplin MA, Anthony R, Dubrowski A. Objective assessment of temporal bone drilling skills. *Ann Otol Rhinol Laryngol* 2007; **116**: 793–798.
- 54 Mackenzie CF, Garofalo E, Puche A, Chen H, Pugh K, Shackelford S *et al.* Performance of vascular exposure and fasciotomy among surgical residents before and after training compared with experts. *JAMA Surg* 2017; **152**: 1–8.
- 55 Mednick Z, Tabanfar R, Alexander A, Simpson S, Baxter S. Creation and validation of a simulator for corneal rust ring removal. *Can J Ophthalmol* 2017; **52**: 447–452.
- 56 Wong K, Stewart F. Competency-based training of basic trainees using human cadavers. *ANZ J Surg* 2004; **74**: 639–642.
- 57 Rowland EB, Kleinert JM. Endoscopic carpal-tunnel release in cadavera. An investigation of the results of twelve surgeons with this training model. *J Bone Joint Surg Am* 1994; **76**: 266–268.
- 58 Levine RL, Kives S, Cathey G, Blinchevsky A, Acland R, Thompson C *et al.* The use of lightly embalmed (fresh tissue) cadavers for resident laparoscopic training. *J Minim Invasive Gynecol* 2006; **13**: 451–456.
- 59 Bergeson RK, Schwend RM, DeLucia T, Silva SR, Smith JE, Avilucea FR. How accurately do novice surgeons place thoracic pedicle screws with the free hand technique? *Spine (Phila Pa 1976)* 2008; **33**: E501–E507.
- 60 Tortolani PJ, Moatz BW, Parks BG, Cunningham BW, Seftor J, Kretzer RM. Cadaver training module for teaching thoracic pedicle screw placement to residents. *Orthopedics* 2013; **36**: e1128–e1133.
- 61 Mowry SE, Hansen MR. Resident participation in cadaveric temporal bone dissection correlates with improved performance on a standardized skill assessment instrument. *Otol Neurotol* 2014; **35**: 77–83.

- 62 Awad Z, Tornari C, Ahmed S, Tolley NS. Construct validity of cadaveric temporal bones for training and assessment in mastoidectomy. *Laryngoscope* 2015; **125**: 2376–2381.
- 63 Egle JP, Malladi SV, Gopinath N, Mittal VK. Simulation training improves resident performance in hand-sewn vascular and bowel anastomoses. *J Surg Educ* 2015; **72**: 291–296.
- 64 Nicandri GT, Cosgarea AJ, Hutchinson MR, Elkousy HA. Assessing improvement in diagnostic knee arthroscopic skill during the AAOS fundamentals of knee and shoulder arthroscopy course for orthopaedic residents. *Orthop J Sports Med* 2015; **3**: 2325967115S0006.
- 65 Kim SC, Fisher JG, Delman KA, Hinman JM, Srinivasan JK. Cadaver-based simulation increases resident confidence, initial exposure to fundamental techniques, and may augment operative autonomy. *J Surg Educ* 2016; **73**: e33–e41.
- 66 Martin KD, Patterson DP, Cameron KL. Arthroscopic training courses improve trainee arthroscopy skills: a simulation-based prospective trial. *Arthroscopy* 2016; **32**: 2228–2232.
- 67 Ciporen JN, Lucke-Wold B, Mendez G, Cameron WE, McCartney S. Endoscopic management of cavernous carotid surgical complications: evaluation of a simulated perfusion model. *World Neurosurg* 2017; **98**: 388–396.
- 68 Martin M, Vashisht B, Frezza E, Ferone T, Lopez B, Pahuja M *et al.* Competency-based instruction in critical invasive skills improves both resident performance and patient safety. *Surgery* 1998; **124**: 313–317.
- 69 Benkhadra M, Faust A, Ladoire S, Trost O, Trouilloud P, Girard C *et al.* Comparison of fresh and Thiel's embalmed cadavers according to the suitability for ultrasound-guided regional anesthesia of the cervical region. *Surg Radiol Anat* 2009; **31**: 531–535.
- 70 Lloyd GM, Maxwell-Armstrong C, Acheson AG. Fresh frozen cadavers: an under-utilized resource in laparoscopic colorectal training in the United Kingdom. *Colorectal Dis* 2011; **13**: e303–e304.
- 71 Eisma R, Lamb C, Soames RW. From formalin to Thiel embalming: what changes? One anatomy department's experiences. *Clin Anat* 2013; **26**: 564–571.
- 72 Giger U, Frésard I, Häfliger A, Bergmann M, Krähenbühl L. Laparoscopic training on Thiel human cadavers: a model to teach advanced laparoscopic procedures. *Surg Endosc* 2008; **22**: 901–906.
- 73 Sullivan GM. Deconstructing quality in education research. *J Grad Med Educ* 2011; **3**: 121–124.

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.

Measuring the educational impact of simulation training in Trauma and Orthopaedics

Hannah James



Hannah James is a Specialist Registrar on the Warwick Rotation, and a member of the BOA Simulation Group. She has done a PhD investigating the impact of cadaveric simulation for T&O surgeons-in-training. Her research interest is in measuring the transfer of surgical skill from the simulated environment to the real-world operating theatre.

Simulation allows trainees to progress from a state of being 'novice' to 'competent' in a highly controllable, standardized environment, remote from patients. Competency based assessment can be used to provide assurance that an individual has reached the required standard following the training, before being allowed to begin performing surgery on real patients.

Before significant investment can be recommended in the widespread provision of simulation for T&O trainees, there needs to be robust evidence that simulation training improves surgical skills in an objectively measurable manner, that skills are translated into the operating theatre, and are retained longitudinally. There is a wide variety of simulation technology available, ranging from inexpensive, low-fidelity bench-top/box-trainers to ultra-high fidelity cadaveric simulation. This raises the important question around the best type of simulation technology to use to achieve the desired educational outcomes in the most timely and cost-efficient manner, and how to time delivery within the training programme.

Measuring impact in simulation research

It is possible to measure the educational impact of a training intervention and this is traditionally described with reference to Kirkpatrick's hierarchy¹ (Fig 1). Evidence

at Level 1 (change in learner reaction) dominates the literature on simulation research, and relies on subjective measures of learner reaction such as confidence and satisfaction. Level 2 studies (change in knowledge) usually rely on written or oral examination assessments to measure learning, and whilst objective, are arguably not appropriate outcome measures for the assessment of technical skill. Level 3 studies (change in learner behaviour) and, in an ideal world, Level 4 (change in patient outcome) provide much more powerful evidence for the effectiveness of simulation training.

A major difficulty with designing Level 3 or 4 studies is finding an appropriate, sufficiently validated outcome measure to test the impact of the training intervention on either learner behaviour or patient outcome. The most commonly used tool to measure behaviour change after simulation training in Level 3 studies is Workplace Based Assessment (WBA), however these are suboptimal in the research setting in view of coarse descriptors, generally low inter-rater reliability and rater effects that compromise validity^{2,3}. Simulator-derived metrics such as procedure time⁴, >>

error frequency⁶, and more recently hand motion analysis⁶ are also frequently used in simulation research. The utility evidence for these assessment tools is promising, but in their relative infancy^{4, 5, 7-11}.

A recent systematic review¹² of simulation in T&O found that of 71 eligible studies, 47 (62%) involved arthroscopy simulators, and that these studies had the highest level of evidence. The evidence base is much more advanced for arthroscopic simulation compared to open surgical simulation, and this may be because of the ready availability of simulator derived metric data for assessment of behaviour change.

'Patient' outcome measures, for Level 4 studies, are equally problematic. Surrogate outcome measures of operative success are often used as an objective assessment of the final product quality and are known as 'final product analysis' (FPA). Examples in T&O simulation research include guidewire position for hip fracture pinning⁴, tip-apex distance for dynamic hip screw¹³, and biomechanical failure point of a fracture fixation construct¹⁴. With the probable exception of tip-apex distance, the utility evidence for FPA methods is generally weak.

The ultimate way of measuring change in patient outcome following simulation training would be to use Patient Reported Outcome Measurements (PROMs) and morbidity/mortality data. The level of 'noise' from confounding variables and difficulty in seeing any effect of the training on patient outcomes using these measures would likely be very significant, and would require a large and well-designed randomised controlled trial to demonstrate proof of effect. Such a study has not (yet) been done.

Challenges of designing simulation research

In addition to the difficulties in finding an appropriate outcome measure, a research team needs to design a sufficiently robust study to answer the research question, using methods that are both appropriate and feasible. Surgical education research is often beset with methodological problems that stem from the real world 'messy' nature

“The ultimate way of measuring change in patient outcome following simulation training would be to use Patient Reported Outcome Measurements (PROMs) and morbidity/mortality data.”



Figure 1: Modified Kirkpatrick's hierarchy.

of education and training. These problems include; small sample sizes (and hence underpowering), lack of randomization and over-use of single-group non-comparative cohort design. There is often reliance on pre-test/post-test assessment strategies which can over-estimate the observed effect size¹⁵. Funding is another barrier, surgical education research does not typically attract significant amounts of research funding and hence precludes delivery of the large scale trials that are needed to provide a high level of evidence for the effectiveness of simulation in training T&O surgeons.

The future of simulation research in T&O

It is encouraging that delivery of simulation in surgical training has been declared a priority area

by the Department of Health¹⁶ and General Medical Council¹⁷, and the development of a simulation mapped curriculum for Trauma and Orthopaedics is well progressed.

A national survey of UK and Republic of Ireland Training Programme Directors is currently being undertaken by the BOA simulation group to map the current provision of simulation within training programmes, and to identify

research activity into simulation. Three doctorates in simulation for T&O by specialist trainees are either recently awarded (one) or in their final stages (two). The future of simulation research in T&O is therefore very promising, and one hopes that as it gains momentum, it will continue to attract more research funding and credibility.

In summary:

- The present evidence base for simulation in T&O training is limited and lacking in high quality evidence showing that skills learnt in the simulated environment are transferred to the operating theatre and retained longitudinally.
- The evidence base for virtual reality arthroscopic simulators is more developed than that of open surgery, low-fidelity box trainers and cadaveric simulation.
- It is difficult to design high quality simulation research studies because of the lack of appropriately validated outcome measures, coupled with pragmatic difficulties including funding.
- Simulation research is increasingly gaining momentum and recognition amongst surgical education stakeholders, and the future looks promising. ■

References

References can be found online at www.boa.ac.uk/publications/JTO.

Use of cadavers to train surgeons: what are the ethical issues?

Hannah James  ^{1,2}

¹Clinical Trials Unit, University of Warwick, Coventry, UK
²Trauma & Orthopaedic Surgery, University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK

Correspondence to

Dr Hannah James, Clinical Trials Unit, University of Warwick, Coventry CV4 7AL, UK; h.smith.1@warwick.ac.uk

Received 21 November 2019
 Accepted 26 November 2019

ABSTRACT

This is an invited submission from the Editor-in-Chief as the introductory piece for an 'Ethics Roundtable'. This piece will include invited commentaries from experts in surgical education, medical ethics, law and the prospective body donor perspective.

BACKGROUND

There is an urgent need to find safe and reliable methods for training novice surgeons. An environment of reduced trainee working hours¹ and increasing service demands at the expense of training opportunities has threatened the quality of postgraduate surgical training.^{2 3} These threats in combination with ethical concerns about the safety implications of novice surgeons 'practising' on live patients⁴ has driven the need to move the early part of the surgical learning curve away from patients and into the simulation laboratory. The traditional master-apprenticeship model of training is no longer fit-for purpose, and a move to a modern, streamlined system of objectively assessed, competency based training is underway.⁵

Human cadavers (deceased bodies) are increasingly recognised as a valuable surgical training resource. Cadavers have excellent potential as a training tool—they have the correct anatomy, soft tissue handling characteristics and neurovascular hazards that would be encountered in the live operating theatre, with zero risk of potential harm to living patients. They offer an ultra high-fidelity simulated operative experience that is unrivalled by bench-models, virtual reality tools and animal models.^{4 6} Cadaveric simulation allows the novice surgeon to learn from mistakes and develop skills without potentially harmful consequences for patients.

The development of dedicated cadaveric training laboratories has led to a substantial increase in the capacity for delivery of cadaveric simulation training⁷ (CST) and the growing demand for CST is evidenced by its considerable popularity among surgical trainees.^{6–8}

HISTORICAL CONTEXT

The use of cadavers for the purposes of teaching anatomy has a long and chequered history.^{9 10} Anatomical study using cadavers is now highly regulated, and there is a well-developed body of opinion as to the appropriate ethical treatment of cadavers in the undergraduate anatomy setting.^{11–13}

The value of cadavers for training surgeons to operate for the benefit of patients was recognised as early as the 18th century when the London

Surgeon-Anatomist Sir Astley Cooper (1768–1841) stated of the surgeon 'He must mangle the living if he has not operated on the dead'.⁹ However, the use of cadavers for postgraduate surgeons-in-training to practice their skills has received much less attention in the ethics literature than their use in anatomy demonstration and dissection.

THE LEGAL POSITION

Recent legislative change has clarified the legal position with regard to using cadavers for surgical training purposes in the UK.

Previously, the 1984 Anatomy Act restricted the use of cadavers to the exploration and definition of topographical anatomy and forbade their use for surgical training.¹⁴ Alongside this, the Human Tissue Act 1961 governed the use of cadaveric tissues for diagnostic, research and postmortem examination purposes but did not give explicit guidance on their use in surgical training. The Human Tissue Act 2004¹⁵ integrated and replaced these Acts and defines five 'licensable sectors': teaching, anatomy, research, public display and the making of a post-mortem examination. Provided that an institution holds the appropriate license ('teaching' in this case) and complies with the mandatory governance requirements, it may legally host surgical teaching activities using cadaveric material. Following enactment of the Human Tissue Act several cadaveric surgical training facilities have developed in the UK associated with an ever-expanding programme of training courses.

ETHICAL CHALLENGES RAISED BY USE OF CADAVERS FOR SURGICAL TRAINING

Unlike anatomical dissection where the body tissues are explored unaltered from their natural state, performing surgery by definition necessitates the removal, rearrangement and sometimes destruction or damaging of tissue or body parts. It may also involve implanting foreign materials, for example, a bone might be deliberately broken to enable a trainee surgeon to practice fixing it with metal pins. This difference in purpose and practice has implications for informed consent and raises questions about the extent to which we are using the body as an object to be manipulated and perhaps destroyed. Cadaveric surgical training courses, in contrast to undergraduate anatomy training, rely heavily on for-profit body donation companies to meet the demand for cadaveric material.¹⁶ The commercialisation of cadavers for surgical training, in addition to the destructive nature of the procedures



© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: James H. *J Med Ethics* Epub ahead of print: [please include Day Month Year]. doi:10.1136/medethics-2019-105873

undertaken, raise ethical considerations that are distinct from those of the anatomical sciences.

The provision of CST is very likely to increase in future. How we treat deceased bodies reflects professional and societal values and attitudes towards the dead and the respect necessarily shown to them because of their human status. Therefore, a discussion of the ethical issues involved with providing cadaveric surgical training is both timely and necessary.

Twitter

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iD

Hannah James <http://orcid.org/0000-0002-0535-3062>

REFERENCES

- Chikwe J, de Souza AC, Pepper JR. No time to train the surgeons. *BMJ* 2004;328(7437):418–9.
- Temple J. *Time for training. A review of the impact of the European working time directive on the quality of training*. London: Medical Education England, 2010.
- Greenaway D. *Securing the future of excellent patient care: final report of the independent review*. London, 2013.
- Reznick RK, MacRae H. Teaching surgical skills — changes in the wind. *N Engl J Med* 2006;355(25):2664–9.
- Improving Surgical Training. Proposal for a pilot surgical training programme 2015.
- Gilbody J, Prasthofer AW, Ho K, et al. The use and effectiveness of cadaveric workshops in higher surgical training: a systematic review. *Annals* 2011;93(5):347–52.
- James H, Chapman A, Pattison G, et al. Current status of cadaveric simulation for surgical training: a systematic review. *British Journal surgery* 2019;106(13):1726–34.
- James H, Thomas S, Pattison G. Paediatric orthopaedic surgery simulation training using short-stature adult cadavers: a feasibility study. *Ann R Coll Surg Engl In Press* 2019.
- Roach M. *Stiff: the curious lives of human cadavers*. London: Penguin Books, 2003.
- Hildebrandt S. Capital punishment and anatomy: history and ethics of an ongoing association. *Clin Anat* 2008;21(1):5–14.
- Cohen N. Is it just semantics? Medical students and their 'first patients'. *J Med Ethics* 2019;45(6):411–4.
- Champney TH. A proposal for a policy on the ethical care and use of cadavers and their tissues. *Anat Sci Educ* 2011;4(1):49–52.
- Champney TH, Hildebrandt S, Gareth Jones D, et al. Bodies R us: ethical views on the Commercialization of the dead in medical education and research. *Anat Sci Educ* 2019;12(3):317–25.
- Mahadevan V. The human tissue act 2004: implications for anatomical work at the College. *Bulletin* 2006;88(8):264–5.
- Human tissue act 2004.
- Champney TH. The business of bodies: ethical perspectives on for-profit body donation companies. *Clin Anat* 2016;29(1):25–9.

Use of cadavers to train surgeons: what are the ethical issues? — body donor perspective

Tracy A Walker,¹ Hannah K James ^{2,3}

In my professional role as anatomy administrator and bequeathal secretary at a large surgical training centre, I am the first point of contact both for people wishing to donate their body, and for newly bereaved relatives telling us that their registered loved-one has died. I am involved in every stage of the process from that first phone call, through to eventual funeral service, cremation of the body and return of the ashes to the family. I am also a registered body donor myself, as I strongly believe in the value of cadaveric training having seen it first hand.

When prospective donors and relatives find out that I am also a registered body donor they find this to be very reassuring to know that having ‘behind the scenes’ access has not put me off, and is a good endorsement for the process, and helps normalise it. I know from first-hand experience that total respect for body donors is central to everything we do here, and often friends/relatives of donors have been so impressed by the bequeathal process that they sign up themselves.

Most donors tell me that the reason they want to donate their body is to be able to give something back to the medical profession and to society as a whole. This is often after they have experienced personal benefit from the National Health Service and surgical treatment, whether that is by curing pain, improving their

quality of life or even saving their life. They feel that knowing that their body could help train many different healthcare professionals after their death, and bring some good that goes beyond their lifetime is the main motivation. They feel that immediately cremating or burying their body after death is a waste.

These are also my main reasons for donating my body, and having recently had some hand surgery that has resulted in a big improvement in my symptoms, I can see the value of having surgeons trained to the best standards using cadavers both from professional and personal perspectives.

Body donation is also important for the families of donors. Newly bereaved relatives, elderly people who have been married for decades and suddenly find themselves widowed, often say to me that respecting their spouse’s wishes by facilitating the donation process is ‘the last thing I can do for him/her’, and ‘I’ll grieve later but I’ve got to get this done for him’. They are at the worst point in their lives, their world has come crashing down, and yet they find comfort in knowing that their loved-ones final wishes have been honoured.

We hold a thanksgiving service for the donors, attended by staff, local dignitaries and students/healthcare professionals who have benefited from the cadaveric training. The donors names are read out at the service and it is always an uplifting and moving experience. Grieving families often tell me that they have great pride and satisfaction at seeing not only how many future medical professionals their loved-one has helped train, but also how appreciated that gift is by those who have benefited.

Most donors do not mind what their body is used for, be that anatomical dissection by medical students to learn anatomy, or for teaching surgeons to do major operations like knee or hip replacements and trauma surgery following a serious accident. I do not see a difference either and would be happy for my body to be used in any of these ways, as all are going towards the greater good of training future healthcare professionals. We do occasionally get asked by body donors or their families to avoid doing certain procedures on the body, and of course we always honour their wishes.

Donating your body is a remarkable gift to the future of healthcare. For me, and for most of our prospective donors, it is easy making the decision to donate as the benefit is so clear. Often, after some media coverage of how fantastic cadaveric training is, we are flooded with new prospective donors!

Twitter

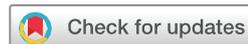
Contributors This piece was written by a prospective body donor (TAW). HKJ provided academic support and wrote the introductory piece for the commentary.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.



To cite Walker TA, James HK. *J Med Ethics* Epub ahead of print: [please include Day Month Year]. doi:10.1136/medethics-2019-105998

Received 28 January 2020
Accepted 2 February 2020



► <http://dx.doi.org/10.1136/medethics-2019-105873>

J Med Ethics 2020;**0**:1.
doi:10.1136/medethics-2019-105998

ORCID iD

Hannah K James <http://orcid.org/0000-0002-0535-3062>

¹Anatomy Administrator & Registered body donor, School of Medicine, Keele University, Staffordshire, UK

²Clinical Trials Unit, University of Warwick, Coventry, UK

³Department of Trauma & Orthopaedic Surgery, University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK

Correspondence to Dr Hannah K James;



**Author response to: Comment on:
Systematic review of the current
status of cadaveric simulation for
surgical training**

Editor

We thank Angelo and Gallagher for their interest in our paper and we congratulate them on their previous important work^{1–4} that contributed to the body of evidence for simulation in surgical training.

The aim of our systematic review was to evaluate the evidence for the impact of *cadaveric* simulation training on surgical trainees. We excluded studies that used cadaveric simulation for the development of proficiency assessment tools^{1,2,4}, as these do not test the educational impact of a cadaveric simulation training intervention. We also excluded studies where cadaveric simulation was used to test the effect of other types of non-cadaveric simulation training protocols. Finally, we excluded studies where it was not clear whether the training intervention used cadaveric simulation. Angelo *et al.*'s study³ was excluded because the intervention group *training* used a plastic model.

We have included their studies in a separate forthcoming systematic review⁵ of technical skill acquisition and operative competency in trauma and orthopaedic surgery.

H. K. James¹ , A. W. Chapman³,
G. T. R. Pattison³, D. R. Griffin² and
J. D. Fisher¹

¹*Clinical Trials Unit and* ²*Warwick Orthopaedics, Warwick Medical School, University of Warwick, and* ³*Department of Trauma and Orthopaedics, University Hospitals Coventry and Warwickshire, Coventry, UK*

DOI: 10.1002/bjs.11501

- 1 Angelo RL, Ryu RKN, Pedowitz RA, Gallagher AG. The Bankart performance metrics combined with a cadaveric shoulder create a precise and accurate assessment tool for measuring surgeon skill. *Arthroscopy* 2015; **31**: 1655–1670.
- 2 Angelo RL, Ryu RKN, Pedowitz RA, Gallagher AG. Metric development for an arthroscopic Bankart procedure: assessment of face and content validity. *Arthroscopy* 2015; **31**: 1430–1440.
- 3 Angelo RL, Ryu RKN, Pedowitz RA, Beach W, Burns J, Dodds J *et al.* A proficiency-based progression training curriculum coupled with a model simulator results in the acquisition of a superior arthroscopic Bankart skill set. *Arthroscopy* 2015; **31**: 1854–1871.
- 4 Gallagher AG, Ryu RKN, Pedowitz RA, Henn P, Angelo RL. Inter-rater reliability for metrics scored in a binary fashion – performance assessment for an arthroscopic Bankart repair. *Arthroscopy* 2018; **34**: 2191–2198.
- 5 James H, Chapman A, Pattison G, Fisher JG, Griffin DR. Assessment of technical skill acquisition and operative competence in trauma and orthopaedic surgical training: a systematic review. *JBJS Rev* 2020; <https://doi.org/10.2106/JBJS.RVW.19.00167>



Correspondence

The Editors welcome topical correspondence from readers relating to articles published in the Journal. Letters should be submitted electronically via the BJS submission site (mc.manuscriptcentral.com/bjs). All correspondence will be reviewed and, if approved, appear in the Journal. Correspondence must be no more than 300 words in length.

Author response to: Comment on: Systematic review of the current status of cadaveric simulation for surgical training

Editor

We thank Jones and Baraza for their interest in our paper¹ and for highlighting the omission of the work of Hughes *et al.*² from the meta-analysis. This was a large piece of work, with more participants than any other paper in the literature. As the educational assessment data were held in a supplementary file, the information was missed and the data were not included in error. We apologize for this.

Here, we present an addendum to Table S2 (supporting information) in our article.

The authors describe delivery of a sophisticated programme of cadaveric training courses using centralized funding for a large number of trainees, which is an outstanding achievement.

We particularly commend the multispeciality and multiprofessional nature of the training. It targets core trainees, the group whom recent qualitative work³ has shown may stand to gain the greatest benefit from cadaveric simulation training.

The purpose of our review was to appraise the methodology and level of evidence of educational impact of studies assessing cadaveric simulation for postgraduate training. Hughes *et al.*'s work² is a non-randomized, non-comparative descriptive study with Kirkpatrick level 1 subjective evidence of impact. WBAs were used to give feedback to participants, although these are not reported in the results. Neither of the outcome measures that are reported have been formally validated for postgraduate training^{4,5}.

This study certainly merits inclusion in the review but does not alter our primary findings. There remains a deficiency of objective evidence of skill transfer to the live operating theatre following cadaveric simulation training. RCTs with clinically relevant outcome measures are required, which we believe is achievable.

H. K. James^{1,2}  and G. T. R. Pattison¹
¹Clinical Trials Unit, Warwick Medical School, and ²Department of Trauma and

Orthopaedic Surgery, University Hospitals Coventry and Warwickshire, Coventry, UK

DOI: 10.1002/bjs.11629

- 1 James HK, Chapman AW, Pattison GTR, Griffin DR, Fisher JD. Systematic review of the current status of cadaveric simulation for surgical training. *Br J Surg* 2019; **106**: 1726–1734.
- 2 Hughes JL, Katsogridakis E, Junaid-Siddiqi A, Pollard JS, Bedford JD, Coe PO *et al.* The development of cadaveric simulation for core surgical trainees. *Bull R Coll Surg Engl* 2018; **101**: 38–43.
- 3 James HK, Pattison GTR, Griffin DR, Fisher JD. How does cadaveric simulation influence learning in orthopedic residents? *J Surg Educ* 2020. <https://doi.org/10.1016/j.jsurg.2019.12.006> [Epub ahead of print].
- 4 Nagraj S, Wall D, Jones E. The development and validation of the mini-surgical theatre educational environment measure. *Med Teach* 2007; **29**: e192–e197.
- 5 Roff S, McAleer S, Harden RM, Al-Qahtani M, Uddin Ahmed A, Deza H *et al.* Development and validation of the Dundee Ready Education Environment Measure (DREEM). *Med Teach* 1997; **19**: 295–299.

Study	Training intervention	Skills taught	Comparator	Primary outcome measure	Results	Skill transfer	MERSQI score
Hughes <i>et al.</i> ² , 2018N = 419 PGY 3–4	10x 2–3-day workshops	Curriculum mapped, not specified	None	DREEM mini-STEEM WBA*	Participants agreeing very useful: as a learning tool, 98%; to improve surgical skills, 82%; to improve confidence 71%	Not measured	5.5

*Not reported. MERSQI, Medical Education Research Quality Instrument; DREEM, Dundee Ready Education Environment Measure; Mini-STEEM, mini-Surgical Theatre Educational Environment Measure; WBA, Workplace Based Assessment.