Imperial College London

Department of Surgery and Cancer

PhD Thesis

Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery

Nathan Curtis BSc BM PGCEd FEBS FRCS

Supervisors

Professor Nader Francis, PhD FRCS

Professor George Hanna, PhD FRCS

December 2019

Declaration

I hereby declare that all work described in this thesis is original and confirm it has been performed by myself. Research collaboration and support from others is explicitly stated in each respective chapter. This thesis has been produced by myself and I confirm that the work presented has not previously been submitted for consideration of any qualification from any other institution.

Nathan Joseph Curtis

London, December 2019

Copyright statement

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence (CC BY-NC-ND 4.0). Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Disclosure statement

The equipment for the 2D3D randomised trial was provided as a free research loan by Karl Storz Endoscopy (UK) Ltd. They had no knowledge or input on the study design, management, performance of the trial, data collection, analysis, this thesis or any outputs arising from it.

Unrelated to this thesis and through open competition I was awarded course and travel fees from Karl Storz UK Ltd to attend the EAES lower GI laparoscopic training course at the Cuschieri Skills Centre, University of Dundee, December 2016.

Karl Storz GmbH (Deutschland) sponsor the trophy and €2500 prize for the "Karl Storz EAES award" at each annual EAES congress. This is awarded to the best oral presentation by a resident in training as determined by the EAES expert panel. I was honoured to receive this prize in Frankfurt, June 2017. Karl Storz held no influence on this decision.

At my request Karl Storz (UK) kindly funded room hire for a regional educational day for higher general surgical trainees in the Wessex Deanery, 15th December 2017, Southampton. In keeping with UK anti-bribery legislation requirements, funding was paid directly to the venue.

Dedications

First and always, to our wonderful daughters Jocelyn and Ottilie. Who came into the world on the first and last days of this undertaking and changed our journey forever.

Most importantly, to my wife Louise. Everything here and beyond is only possible with you. This work required far too many sacrifices from you, and I remain in awe.

To the patients that believed in this work and sought to help those who will face their difficult situation in the future. Humbling, inspiring and dedicated with the greatest respect.

Acknowledgements

This was no small undertaking and I'm hugely grateful to all who have helped me along this road.

To my supervisors, Nader and George, I owe a huge amount. Their immense expertise and relentless focus on providing the best for our patients through their considerable research efforts remains inspiring. I have benefitted for their support, clarity, trust and generosity with their invaluable time.

I remain thankful for the financial support provided by the Board of Trustees and the Clinical Research Unit, Yeovil District Hospital NHS Foundation Trust. Their belief and investment in me at a time of considerable pressure is humbling. I gratefully acknowledge the grant support from the EAES research committee and the Imperial Dean's Award that also made this research possible.

All at the clinical research unit at Yeovil District Hospital particularly Joanna Allison, Nicky Marks and Lucy Pippard who have assisted every step in the process and guided me through the regulatory maze required to perform and successfully deliver multi-centre and multi-national (Wales!) research.

I owe much to my predecessor, Mr Jake Foster (Imperial College London & Yeovil) who kindly gave his time and patience to help train me in the complexities of surgical video analysis and tool use.

Multi-centre research is only possible with the kind assistance of a huge number of skilled people. I'm hugely appreciative of the ten surgeons, expertly backed by their research teams, whose participation added to their considerable workloads and offered up for their performances to intense scrutiny! My operative training has been greatly advanced by these opportunities.

My thanks to all at Karl Storz UK particularly Richard Atkinson, Paul Lewis, Lewis Thorpe and Fred Dale. The trial would not have been possible without their backing and logistical support.

Finally, and most importantly, I am still touched by the enthusiasm and dedication of all the patients who chose to enter the trial. At times of great personal difficulty they universally showed desire to help those who will find themselves in the same position.

Abstract

Introduction: The role of laparoscopy in the surgical management of rectal cancer is debated. Randomised trials have reported contrasting results with inadequate specimens obtained in a minority of patients. The reasons behind these findings are unclear. Complex surgical interventions and human performance are prone to variation, which may account for outcome differences, but neither are robustly measured. Application of quality assurance (QA) to the intraoperative period could explore surgical performance and any relationship with subsequent outcomes. The overarching aim of this thesis is the promotion of oncological and patient safety through application of QA to laparoscopic TME surgery.

Methods: Evidence synthesis of QA tools was obtained through a systematic review to identify reported objective laparoscopic total mesorectal excision (TME) assessment tools. Development of novel QA tools for laparoscopic TME was performed and applied and validated using case video from two multicentre randomised trials with reliability and validity of the laparoscopic TME performance tool (L-TMEpt) assessed. A multicentre randomised trial comparing 3D vs. 2D laparoscopic TME was performed incorporating objective performance analyses. Scores divided surgeons into quartiles and compared with histopathological and clinical endpoints. A novel intraoperative adverse event classification was developed and piloted.

Results: 176 cases from 48 credentialed surgeons were analysed. L-TMEpt inter-rater, test-retest and internal consistency reliabilities were established. Substantial variation in surgical performance were seen. Scores were strongly associated with the number of intraoperative errors, plane of mesorectal dissection and short-term patient morbidity. Upper quartile surgeons obtained excellent results compared with the lower quartile (mesorectal fascia 93% vs. 59%, NNT 2.9, p=0.002; 30-day morbidity 23% vs. 48%, NNT 4, p=0.043).

Conclusions: Intraoperative QA using assessment tools can objectively and reliably measure complex cancer interventions. Laparoscopic TME surgical performance assessment showed substantial variation which is strongly associated with clinical outcomes holding implications for surgical trial design and interpretation.

Quotations

Are doctors who make mistakes villains? No, because then we all are.

- Better: A Surgeon's Notes on Performance by Atul Gawande.

I realised it's not about me, it's about the patient. I need to say, 'I might be the reason, fix me'.

- Dr Henry Otero, Consultant Oncologist, reflecting on a personal error identified in a root cause analysis following a clinical incident. Quoted from Black Box Thinking by Matthew Syed.

We learn from our mistakes. It is as simple and as difficult as that.

- Garry Kaplan, Chief Executive Officer, Virginia Mason Hospital, Seattle, WA, USA.

  On receiving the Distinguished Hospital Award for Clinical Excellence for the 8th successive year.

Table of contents

Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery 1

Declaration 2

Copyright statement 2

Disclosure statement 3

Dedications 4

Acknowledgements 5

Abstract 6

Quotations 8

Table of contents 9

Outputs from this thesis 16

List of abbreviations used in this thesis 19

List of Tables 21

List of Figures 23

Chapter 1 – Introduction, background and the aims of this thesis 25

1.1 The anatomy of the rectum 26

1.2 Rectal cancer 28

1.3 Surgical resection of the rectum 32

Previous problems and the evolution of rectal cancer management 33

Staging 33

Neoadjuvant treatment 34

MDT working practices 35

1.4 Surgical Management - Anterior resection of the rectum 35

Total mesorectal excision – procedural steps 35

TME and patient outcomes 37

Quality of life and pelvic nerves 38

1.5 Minimal access surgery 38

Laparoscopic colon cancer resection 39

1.6 Minimal access surgery for rectal cancer 40

Technological developments 41

Definition of laparoscopic TME 41

Large randomised trials and current controversies 41

1.7 Surgical performance and patient outcome in rectal cancer surgery 46

1.8 Assessment theory applied to surgery 48

1.9 Transition in surgical assessment 49

Assessment tools 50
Assessment of specialists 52
Implications of specialist assessment 53
Validity and Reliability 53
1.10 Example of specialist assessment in laparoscopic surgery 55
1.11 Assessing surgery 58
1.12 Quality assurance of operative performance 60
1.13 Application of quality assurance and quality control to colorectal surgery 61
1.14 Discussion 65
1.15 Conclusions and rationale for this thesis 67
1.16 Hypotheses and objectives of this thesis 68
Hypotheses 68
Objectives 68
Chapter 2 – Evidence synthesis on the objective assessment of laparoscopic total mesorectal surgery 69
2.1 Evidence Synthesis - Introduction 70
Aims 70
2.2 Methods 70
Inclusion Criteria 73
Data extraction 73
2.3 Results 74
Objective intraoperative TME assessment tools 74
Types of the assessment tools and methods 78
Aims of the studies and purpose of the tools 78
Formative assessment tools 78
Global assessment scale 78
Global modular attainment score 79
Performance-based assessment (PBA) 79
Summative assessment tools 79
Laparoscopic competency assessment tool (L-CAT) 79
Hierarchical task analysis (HTA) 80
Research assessment tools 80
Observational clinical human reliability analysis (OCHRA) 80
Operative nerve injury assessment 81
Credentialing of learners and assessors 81
Validity and reliability of the tools 82

2.4 Application of objective assessment within multicentre laparoscopic TME randomised trials 84
2.5 Critical appraisal of results and the identified tools 85
2.6 Addressing the identified need 88
Features of an ideal TME assessment tool 88
The laparoscopic TME performance tool (unpublished work) 88
Development of the L-TMEpt 90
2.7 Selection of tools for application to rectal cancer trials 92
2.8 Conclusions 93
Chapter 3 – Testing the application of intraoperative quality assurance to a laparoscopic total mesorectal excision randomised trial 94
3.1 Introduction 95
Aim 95
3.2 Trial selection for surgical performance assessment 95
3.3 ALaCaRT trial synopsis 98
3.4 Contact and sub study proposals and approvals 99
3.5 Methods 99
Objectives 99
OCHRA training 100
Reliability evaluation 100
Concurrent validity evaluation 101
Development of OCHRA methodology for laparoscopic TME 101
L-TMEpt assessment tool 104
Establish a L-TMEpt score (threshold) cut off 104
Statistical analysis 104
Data governance 105
3.6 Video series and application of the laparoscopic TME performance tool 105
Revised statistical analysis plan 107
3.7 Results 108
OCHRA training 108
Analysis Time 108
OCHRA data 108
Laparoscopic TME performance tool results 113
L-TMEpt inter-rater reliability 115
Inter-item reliability 115
L-TMEpt test-retest reliability 116
L-TMEpt case complexity test retest reliability 119

L-TMEpt inter-item reliability assessment 119
OCHRA and L-TMEpt concurrent validity 120
3.8 Clinical validity of objective assessment methods 121
Histopathological composite endpoint 122
30 day morbidity 124
3.9 Discussion 126
3.10 Conclusion 130
Chapter 4 – Prospective application of intraoperative quality assurance within a laparoscopic total mesorectal excision trial 131
4.1 Copyright statement 132
4.2 Introduction 132
Aim 132
4.3 Laparoscopic TME case video searches 132
4.4 Conception and study planning 133
4.5 Trial clinical research justification – 3D laparoscopy 134
4.6 Methodology and study design 136
Strategy 136
The application of intraoperative QA as study endpoints 136
2D/3D trial hypothesis 137
Sample size 137
Timeline 138
Patient inclusion criteria 140
Surgeon inclusion criteria 140
Surgeon stereopsis testing 140
Recruitment and randomisation protocol 141
Laparoscopic equipment set up 141
Intervention standardisation 141
Data Collection 142
OCHRA methodology 143
Pelvimetry measurements 143
Statistical analysis 145
4.7 Regulatory and funding approvals 145
4.8 Trial delivery 146
Substantial amendment 146
4.9 Results 147
Recruiting surgeon and stereopsis testing 147

Recruitment and participant feedback 147
Trial CONSORT diagram and patient attrition 148
Patient and tumour demographics 149
4.10 Application of objective assessment tools to laparoscopic TME performance 151
OCHRA data 151
OCHRA error analysis 153
Laparoscopic TME performance tool data 156
Content validity – case complexity 158
Validity of TME objective assessment 160
Concurrent Validity 160
Clinical validity 162
30 day morbidity 164
4.11 Prospective application of intraoperative quality assurance to 2D vs. 3D surgery 168
4.12 2D vs. 3D laparoscopic TME RCT clinical results 170
Operative data 170
Case complexity – Surgeon reported and cognitive load 171
30 day morbidity 173
4.13 Informing the design of future clinical 2D vs. 3D RCT 177
Power calculation for a definitive 3D laparoscopic TME trial 179
4.14 Discussion 180
4.15 Critique and study limitations 185
4.16 Conclusions 188
Chapter 5 – Can quality assurance measure performance variability and associated clinical impacts? 190
5.1 Introduction 191
Aim 192
5.2 Methods 192
Combining ALaCaRT and 2D3D datasets 192
Combined video analysis 193
Clinical, construct and concurrent validity analyses 193
Statistical analysis and data handling 194
5.3 Results 195
Clinical Validity 195
Concurrent Validity 197
Construct Validity 198
Oncological outcomes 200

Case score quartiles (TMEdt) 201
Full L-TMEpt score data 201
5.4 Discussion 202
Variation 203
Comparison of results with the laparoscopic TME evidence base 204
Implications for surgical intervention studies 204
Limitations and areas for future study 205
5.5 Conclusion 208
Chapter 6 – Development and validation of an intraoperative adverse event and near misses reporting classification 209
6.1 Copyright statement 210
6.2 Introduction 210
Aim 211
Chapter objectives 212
6.3 Methods 212
Conception and expert surgeon questionnaire design 212
Classification development 213
Second questionnaire phase 214
Clinical validity 215
Reliability assessments 216
Statistical analysis 216
6.4 Results 217
Expert surgeon participants 217
Survey responses 217
Ideal characteristics 217
Intraoperative adverse event classification 219
6.5 Reliability of the EAES classification 222
6.7 Clinical validation of the intraoperative adverse events and near miss classification 223
2D3D cases - highest grade event per case 223
2D3D cases - sum of non-consequential events 223
Combined ALaCaRT and 2D3D pelvic tasks - highest grade event per case 225
Combined ALaCART and 2D3D pelvic tasks - sum of non-consequential events 226
6.7 Discussion 231
Near miss (non-consequential) events 233
6.8 Limitations and future research 235
6.9 Conclusions 239

Chapter 7 – Discussion and Conclusions 240
7.1 Background synopsis and summary of chapters 241
7.2 Discussion of hypothesis one: Quality assurance tools for laparoscopic total mesorectal excision can be developed and applied to research and clinical settings 243
7.3 Discussion of hypotheses two: Quality assurance tools can reliably assess operative technique and correlate with clinical outcomes following laparoscopic total mesorectal excision 244
7.4 Discussion of specific objectives 245
Identify the reported tools for objective assessment of intraoperative laparoscopic TME performance 245
Apply objective assessment tools to randomised controlled trials to evaluate: 245
Tool reliability 245
Concurrent validity 245
Construct and clinical outcome validity 245
7.5 Implications for the research setting and future research on this topic 246
7.6 Implications for laparoscopic TME clinical practice 250
7.7 Limitations of this thesis 252
7.8 Personal reflections 256
7.9 Concluding remarks 256
Reference list 258
Appendix 1 – Permissions for reproduction of previously published figures and tables 273
Appendix 2 – TMEdt individual case score quartile analysis 276
Appendix 3 – Full laparoscopic TME performance tool data 278

Outputs from this thesis

The following peer-reviewed publications, presentations to learned societies and academic prizes arose from the original work in this thesis:

Peer-reviewed publications

Chapter 2

Objective assessment of minimally invasive total mesorectal excision performance: a systematic review. Curtis NJ, Davids J, Foster JD and Francis NK. Techniques in Coloproctology. 2017 Apr;21(4):259-268. doi: 10.1007/s10151-017-1614-z.

Chapter 4

2D vs. 3D laparoscopic total mesorectal excision: a multicentre randomised controlled trial. Curtis NJ, Conti JA, Dalton R, Rockall TA, Allison AS, Ockrim JB, Jourdan IC, Torkington J, Phillips S, Allison J, Hanna GB & Francis NK. Surgical Endoscopy 2019 Jan 17. doi: 10.1007/s00464-018-06630-9

The use of 3D laparoscopic imaging systems in surgery. EAES consensus development conference 2018. Arezzo A, Vettoretto N, Francis NK, Bonino MA, Curtis NJ, Amparore D, Arolfo S, Barberio M, Boni L, Brodie R, Bouvy N, Cassinotti E, Carus T, Checcucci E, Custers P, Diana M, Jansen M, Jaspers J, Marom G, Momose K, Müller-Stich BP, Nakajima K, Nickel F, Perretta S, Porpiglia F, Sánchez-Margallo F, Sánchez-Margallo JA, Schijven M, Silecchia G, Passera R, Mintz Y. Surgical Endoscopy 2018 Dec 4. doi: 10.1007/s00464-018-06612-x.

Factors predicting operative difficulty of laparoscopic total mesorectal excision. Curtis NJ, Thomas C, Dennison G, Ockrim JB, Conti JA, Dalton R, Allison AS & Francis NK. Diseases of the Colon & Rectum 2019 Dec;62(12):1467-1476. doi: 10.1097/DCR.0000000000001490

Chapter 5

Surgical skill assessment determines clinical outcomes in cancer surgery. Curtis NJ, Foster JD, Miskovic D, Brown CSB, Hewett PJ, Abbott S, Hanna GB, Stevenson ARL & Francis NK. JAMA Surgery 2020 May 6. doi: 10.1001/jamasurg.2020.1004.

Chapter 6

EAES classification of intraoperative adverse events in laparoscopic surgery. Francis NK, Curtis NJ, Conti JA, Foster JD, Bonjer HJ and Hanna GB on behalf of the EAES committees. Surgical Endoscopy, 2018 Sep;32(9):3822-3829. doi: 10.1007/s00464-018-6108-1

Clinical evaluation of intraoperative near misses in laparoscopic rectal cancer surgery. Curtis NJ, Dennison G, Brown CSB, Hewett PJ, Hanna GB, Stevenson ARL and Francis NK. Ann Surg. 2019 Jul 3. doi: 10.1097/SLA.0000000000003452.

Oral Presentations

Classification of intraoperative errors in laparoscopic colorectal surgery. 25th congress of the EAES, Frankfurt-Am-Main, Germany, June 2017.

Adverse events in laparoscopic total mesorectal surgery – what, where, when and why? Invited oral presentation. 26th EAES congress, London, June 2018.

The role of 3D laparoscopy in rectal cancer resection. Invited oral presentation. Spring Colorectal meeting of the Colorectal Surgical Society of Australia and New Zealand, Freemantle, Western Australia, October 2018.

Surgical performance and outcomes following laparoscopic rectal cancer resection. Invited presentation at the St Mark's Hospital London, Grand Round, May 2019.

Validity and reliability evaluation of the EAES intraoperative adverse events classification for minimal access surgery. Gerhard Buess award session, 27th EAES congress, Seville, Spain, June 2019.

Near misses in minimal access surgery: common and under reported. 27th EAES congress, Seville, Spain, June 2019.

Surgical skill assessment determines clinical outcomes in cancer surgery. Colorectal Tripartite Meeting, Auckland, New Zealand, November 2020 (conference postponed to 2022).

Poster Presentations

Does pelvimetry predict intraoperative laparoscopic total mesorectal excision performance?

Poster presentation to the annual meeting of the Association of the Coloproctology of Great Britain & Ireland, Birmingham, July 2018.

3D laparoscopy in total mesorectal excision: a multicentre randomised controlled trial.

Poster presentations to both the European Society of Coloproctology, Nice, France, September 2018 and the Association of Coloproctology of Great Britain & Ireland, Birmingham, July 2018.

Prizes awarded

Cancer research excellence in surgical trials (CREST) UK national award 2017 – National Institute of Health Research (£1000).

Best oral presentation at the 25th congress of the EAES, Frankfurt-Am-Main, Germany, June 2017. Karl Storz & EAES prize (€2500).

Dean's Internship Award, April 2018 – Faculty of Medicine, Imperial College London (£3000).

List of abbreviations used in this thesis

2D	Two dimensional
3D	Three dimensional
2D/3D	The 2D3D laparoscopic TME randomised controlled trial
ALaCaRT	Australasian Laparoscopic Cancer of the Rectum trial
ANZCTR	Australia and New Zealand Clinical Trials Registry
APER	Abdominoperineal excision of the rectum
AUROC	Area under the receiver operator curve
CONSORT	Consolidated Standards of Reporting Trials
CRM	Circumferential resection margin
CRN	Clinical research network
CT	Computerised tomography scan
EAES	European Association of Endoscopic Surgeons
GAS	Global assessment scale
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HRA	UK NHS Health Research Authority
HTA	Hierarchical task analysis
IDEAL	Idea or Innovation, Development, Exploration, Assessment and Long-term Study
ICC	Intraclass correlation coefficient
IMA	Inferior mesenteric artery
IMV	Inferior mesenteric vein
IQR	Inter quartile range
IRAS	Integrated research application system
ISRCTN	International Standard Randomised Controlled Trial Number Registry
LapCo	UK national training programme in laparoscopic colorectal surgery

LoREC	English national Low Rectal Cancer development programme
L-CAT	Laparoscopic competency assessment tool
L-TMEpt	Laparoscopic TME performance tool
MAS	Minimal Access Surgery
MDT	Multi-disciplinary team
MRI	Magnetic resonance Imaging
NASA-TLX	National air and science administration – Task load Index
NHS	National Health Service
NICE	National Institute for Clinical Excellence
NIHR	UK National Institute for Health Research
NNT	Number needed to treat
OCHRA	Observational Human Clinical Reliability Analysis
OSATS	Objective structured assessment of technical skill
PBA	Performance based assessment
PCR	Pathological complete response (to neoadjuvant chemoradiotherapy)
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QA	Quality Assurance
RCT	Randomised controlled trial
SD	Standard deviation
SMA	Superior mesenteric artery
SPSS®	Statistical package for social sciences
STARRCAT	Surgical timing after radiotherapy for rectal cancer trial
TME	Total mesorectal excision
TMEdt	Sum of the first three L-TMEpt TME dissection columns
TNM	TumourNodeMetastasis staging classification (5^{th} edition unless stated)

List of Tables

Table	Page	Description
1.1	31	The 5th and 8th editions of the TNM classification for colorectal cancer
1.2	45	The laparoscopic vs open TME RCTs and meta-analyses
1.3	54	Individual components of validity assessment and testing
1.4	54	The difference aspects of test reliability
1.5	63-4	Methods suggested for quality assurance in multicentre trials involving laparoscopic colorectal surgery
2.1	72	The literature search string
2.2	76-7	PRISMA checklist
2.3	82-3	Details of each included study in the systematic review
2.4	85	The six laparoscopic vs. open RCTs that included attempts at assessment
3.1	96	The laparoscopic TME RCTs that recorded case video
3.2	98	Synopsis of the ALaCaRT trial
3.3	103	Laparoscopic TME hierarchical task analysis
3.4	107	ALaCaRT study IDs that underwent objective performance assessment
3.5	109	Errors per hierarchical task phase
3.6	111-2	OCHRA category lists and ALaCaRT results
3.7	113	Pelvic location of OCHRA errors from ALaCaRT cases
3.8	114	L-TMEpt raw score data for the ALaCaRT cases
3.9	115	Inter-rater L-TMEpt scores
3.10	115	Inter-rater case complexity grades
3.11	116	Inter-rater internal consistency for each L-TMEpt task domain
3.12	119	Test-retest case complexity grading
3.13	119	Test-retest internal consistency for each L-TMEpt task domain
3.14	121	ALaCaRT and OCHRA ALaCaRT data
3.15	123	Objective assessment scores for cases meeting the ALaCaRT endpoint
3.16	124	Assessment scores for CRM and mesorectal fascial plane cases
3.17	125	Objective assessment scores and development of 30 day morbidity
4.1	143	Definitions for TME dissection grade assessment
4.2	145	Pelvimetry measurements
4.3	146	Interim 2D/3D trial TME dissection plane assessment
4.4	147	Interim 2D/3D trial readmission data

4.5	150	2D3D trial patient demographics and tumour characteristics
4.6	155-6	OCHRA category lists and 2D/3D trial arm data
4.7	157	L-TMEpt item, column, row and totals for each trial arm
4.8	159	Pelvimetry data for each case complexity grade
4.9	159	Pelvimetry data and gender
4.10	159	L-TMEpt data for each complexity grade
4.11	163	L-TMEpt data by plane of mesorectal dissection
4.12	167	L-TMEpt data for patients with and without any post-operative morbidity
4.13	169	L-TMEpt data for each 2D/3D trial arm
4.14	172	Surgeon reported case and task difficulties
4.15	174	Number of 30 day morbidity events per patient
4.16	175	Clavien-Dindo grade of 30 day morbidity events for the trial arms
4.17	176	Pathological data for the trial arms
4.18	179	Sample size calculations for a future definitive 2D/3D TME RCT
5.1	194	Cohort demographic data
5.2	195	L-TMEpt data for the combined trial analyses
5.3	198-9	Study data displayed by surgeon quartiles
5.4	200	Oncological outcome data for the ALaCaRT patients
6.1	219	EAES intraoperative adverse event classification with pilot data
6.2	221	Inter-rater EAES grade reliability
6.3	221	Test-retest EAES grade reliability
6.4	222	Crosstabulation of highest EAES grade and morbidity
6.5	223	Near misses and 30 day morbidity
6.6	224	Crosstabulation of EAES grades and morbidity
6.7	225	EAES grade distribution and number of morbidities and severity
6.8	227	EAES intraoperative adverse event classification with combined pelvic data
6.9	228	Near miss event number and clinical outcomes
6.10	229	Pelvic near misses and pathological outcomes

List of Figures

Figure	Page	Description
1.1	27	The anatomy of the rectum. Arterial supply and mesorectum cross section
1.2	32	Histopathological examples of resected rectal cancers
1.3	34	MRI rectal cancer images
1.4	36	The original TME description and resulting overall survival
1.5	36	MRI slice showing direct extension of a rectal cancer into the mesorectum
1.6	47	Oncological outcomes for differing TME specimen qualities
1.7	51	Performance based assessment form for anterior resection of the rectum
1.8	57	The competency assessment tool for laparoscopic colectomy (L-CAT)
2.1	75	PRISMA flow diagram for the systematic review
2.2	91	The laparoscopic TME performance tool (L-TMEpt)
3.1	104	Pelvic spatial zones for defining error locations
3.2	109	OCHRA error count distribution and frequency
3.3	113	ALaCaRT TMEdt score distribution and frequencies
3.4	114	Box whisker plot containing TMEdt by case complexities
3.5	116	Test-retest ALaCaRT TMEdt scatterplot
3.6	118	Bland-Altman plot for ALaCART TMEdt test retest scores
3.7	120	Concurrent validity OCHRA and TMEdt ALaCaRT scatterplot
3.8	122	ALaCaRT TMEdt and body mass index scatterplot
3.9	122	TMEdt area under the receiver operator curve for ALaCaRT cases
4.1	139	2D/3D trial Gantt chart
4.2	148	2D/3D trial recruitment
4.3	149	2D/3D CONSORT flow diagram
4.4	151	Sequential trial case error counts
4.5	152	Scatterplot exploring temporal differences in assessment
4.6	152	OCHRA error distribution and frequency
4.7	153	Distribution of OCHRA errors between hierarchical task phases
4.8	157	L-TMEpt case scores
4.9	157	L-TMEpt scores obtained during the duration of the trial and analyses
4.10	160	Box-whisker plot for L-TMEpt scores and case complexity grades
4.11	160	Scatterplot depicting OCHRA and L-TMEpt data for the 2D/3D trial
4.12	161	OCHRA and L-TMEpt scatterplot with grade complexity overlay

4.13	162	Dual axes scatterplot for L-TMEpt, OCHRA and operative time
4.14	163	Box-whisker plot for L-TMEpt and OCHRA scores and TME dissection plane
4.15	164	AUROC for objective assessment in the 2D/3D trial
4.16	165	Box-whisker plot for OCHRA data and 30-day morbidity outcomes
4.17	166	Box-whisker plot for OCHRA and highest Clavien-Dindo grade per patient
4.18	168	AUROC for objective assessment and morbidity outcomes
4.19	169	Box-whisker plots for 2D/3D trial arm objective assessment data
4.20	170	Scatterplot depicting OCHRA and L-TMEpt data with trial arm overlay
4.21	173	NASA task load index data for the trial arms
4.22	178	Bar chart displaying TME plane of dissection for 2D and 3D surgery
5.1	195	TMEdt score distribution for the combined trial cases
5.2	196	Box and whisker plot showing TMEdt between different case complexities
5.3	196	Scatterplot showing OCHRA and TMEdt data from both trials
6.1	217	Expert ranking of the ideal reporting characteristics
6.2	220	Expert responses to the draft classification
6.3	223	EAES grade I adverse event numbers and distribution
6.4	226	Pelvic near miss distribution
6.5	233	Heinrich's law of industrial safety
6.6	233	The surgical safety pyramids

Chapter 1 – Introduction, background and the aims of this thesis

This thesis investigates the application of quality assurance processes to the intraoperative performance of laparoscopic total mesorectal excision for rectal cancer. This chapter explores rectal anatomy, rectal cancer and the surgical challenges it presents. The need for high quality surgery is reported through the established links with patient outcomes. Assessment theory and its application within to surgery is discussed. Quality assurance is presented alongside its present uses in surgery and areas of potential expansion. The contemporary clinical literature is presented to shape and define the aims and objectives of this thesis.

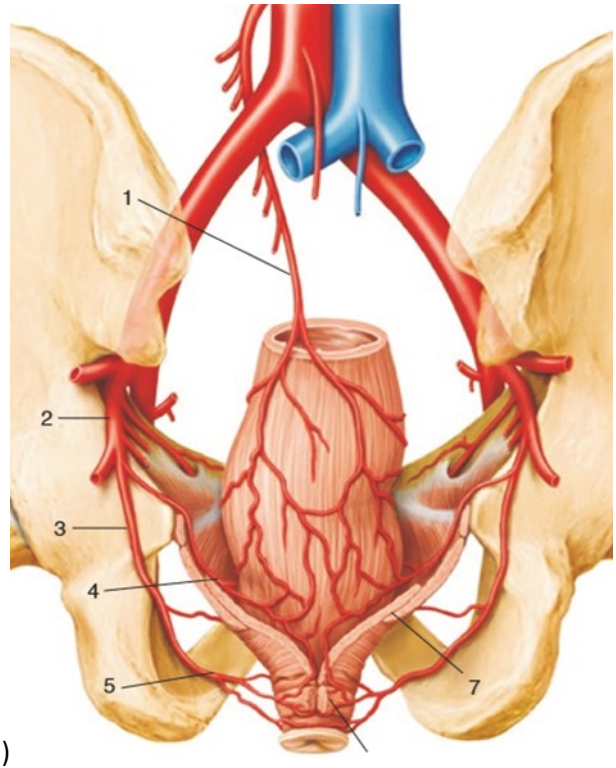
I gratefully acknowledge Dr Cen Thomas (Consultant Radiologist) and Dr Ed Cooper (Consultant Histopathologist, both Yeovil District Hospital NHS Foundation Trust) for supplying some of the original figures used in this chapter.

1.1 The anatomy of the rectum

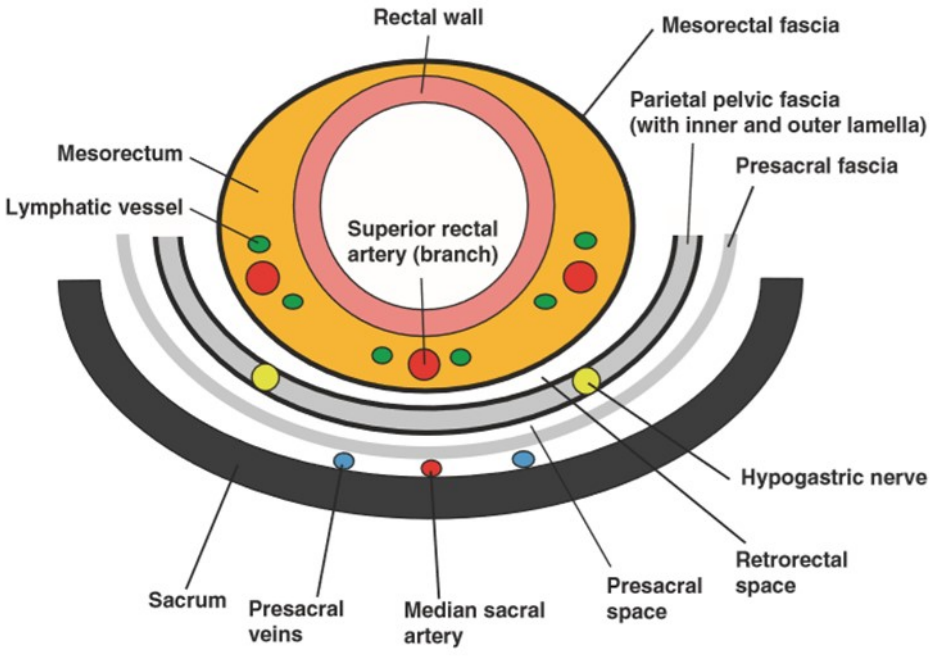
The rectum is the distalmost portion of the alimentary tract. Situated in the posterior compartment of the pelvis it originates in continuity with the sigmoid colon and continues caudally to reach the pelvic floor and anal sphincter complexes. When arbitrarily divided into thirds, the anterior and lateral aspects of the superior rectum are covered by peritoneum. Progressing distally, only the anterior aspect of the middle rectum is coated with all of the distal rectum within the pelvis.

Surprisingly there remains no widely accepted definition of the rectum (1) which has led to a number of variations entering regular clinical use (2). Anatomically, the rectum can be said to begin at the convergence of the three teniae coli (longitudinal muscle bands) at the rectosigmoid junction reported to correspond with the S3 vertebral body level. A number of countries, including the UK, use a fixed distance from the anal verge to define the proximal extent of the rectum but this may oversimplify the natural anatomical variations seen in the population (3). Occasionally a landmark such as the sacral promontory or anterior peritoneal reflection is used which may represent a more patient centred method, but the location of these points can also vary (4). In keeping with contemporary UK surgical and multidisciplinary practice, in this thesis the upper border of the rectum is defined as 15cm proximal to the anal verge.

The distalmost aspect of the rectum is easier to define anatomically as the dentate line, where the embryological endoderm of the hindgut meets the squamous mucosa of the proctoderm (5). From a practical consideration, surgeons often consider the rectum to end at the point where the bowel wall passes through the musculature of the pelvic floor as when viewed from an abdominal approach represents the usual distal extent of procedures (6). This too is being challenged by the increasing utilisation of trans-anal surgical approaches to the rectum (7).



a)



b)

Figure 1.1: a) Arterial supply to the rectum and anal canal viewed posteriorly. 1 – The superior rectal artery (from the IMA). 4 – middle rectal artery and 5 – inferior rectal artery arise from the internal iliac system. b) Cross sectional illustration of the mesorectum, its contents and relationship with surrounding structures. Both figures reproduced from Wedel (8) with permission (Appendix 1).

In keeping with its embryological origins, the rectum receives dual arterial supplies. The inferior mesenteric artery arises anteriorly from the aorta at the level of the L3 body. Typically, after a short 2-3cm course it gives off a left ascending colic branch before branching into sigmoid arteries. The superior rectal artery travels caudally in the sigmoid mesentery to the supply the proximal rectum. Here, encased within the mesorectum, it forms a rich anastomosis with the bilateral middle and inferior rectal arteries that arise from the internal iliac systems. The venous drainage follows a similar distribution with the proximal rectum draining to the hepatic portal system via the superior rectal and inferior mesenteric veins. The distal rectum drains via the internal iliac vein system.

As per the rest of the small and large bowel, the rectal mesentery (mesorectum) contains vasculature and lymphatic tissues. In contrast to the proximal gut, the mesorectum forms a cylinder around the rectal muscular tube. Its outer border is a layer of fascia that envelopes the mesorectum. This mesorectal fascia also acts a support structure which helps holds the anatomical position whilst allowing distention in response to its contents. The mesorectum is surrounded by fixed structures; sacrum posteriorly, pelvic side walls laterally and the prostate or vagina anteriorly (6).

1.2 Rectal cancer

Cancer arising from the rectum is a common malignancy and presents a number of challenges to multi-disciplinary colorectal clinicians. Of the 30710 colorectal cancers diagnosed in England and Wales between 1st April 2015 to 31st March 2016, 8592 (28%) were located in the rectum representing the most frequent anatomical site for bowel cancer (9).

Over 90% of malignancies arising in the rectum originate from the glandular columnar epithelium with their neoplasia termed adenocarcinoma. Rarer malignancies include neuroendocrine cancers (previously called carcinoid), primary lymphomas or connective tissue-based pathologies but given their rarity, in this thesis the term rectal cancer is applied solely to adenocarcinoma. Normal mucosa cells undergo a series of genetic mutations to key tumour suppressor and oncogenes in a sequence

known as the adenoma-carcinoma pathway (10, 11). Progression in the understanding of the differing molecular science and pathophysiological changes that can occur have led to four distinct colorectal cancer subtypes based upon tumour biology being defined by international consensus (12). Correlation with patient outcomes has been shown and personalisation of care based upon individual colorectal cancer biology may lead to a radical shift in the staging, management and prognosis for future patients (12). As this data does not presently influence routine surgical practice or standard of care for patients on curative pathways, this thesis groups all rectal cancers together. It is hoped this oversimplification is addressed in coming times.

In keeping with the behaviour of most solid cancers, rectal cancer can spread in four ways, all with clinical and surgical significance (13). Direct invasion into and subsequently through the muscularis propria and without treatment may continue through the serosal layer or into adjacent viscera. Direct proximal and distal longitudinal and circumferential luminal spread also can occur. Haematological spread, where viable tumour cells gain access to the circulation through vessel wall invasion risking spread both along their course (intra/extra mural vascular invasion) and to distant sites. In keeping with the venous drainage of the rectum the liver represents the most common site for colorectal cancer metastases. Lymphatic channels and nodes draining the rectum are encased in the mesorectum and in keeping with colonic anatomy follow the course of the arterial vessels. Where a proximal or middle rectal cancer breaches the encasing mesorectal fascia and visceral peritoneum peritoneal metastases can result from trans-coelomic dissemination.

Reflecting these processes, the recent 8th revision of the TNM staging system was adopted by the UK Royal College of Pathologists in January 2018 (14). This contained a number of updates in response to increasing understanding and clinical implication of histologically assessed tumours as well as evolution of multi-disciplinary care allowing expansion of indications for treating metastatic disease. The primary differences relate to switching of the pT4a/b nomenclature, formally recognising tumour deposits and their implications rather than assuming they represented replaced lymph

nodes and an expansion of pM1 given the increase in curative treatment offered for metastatic colorectal adenocarcinoma since publication of the TNM 5th edition in 1997 (15).

As the majority of experiments designed and reported in the thesis and cited literature was published prior to January 2018, unless otherwise stated the TNM 5th edition is used throughout. The 2010 TNM 7th edition was not formally adopted in the UK due to concerns regarding interpretation of the evidence used to create the revision, inter-observer variability (16) and negligible difference in the prognostic value over the 5th edition (17).

TNM 5th edition		TNM 8th edition	
Primary Tumour (pT)			
Tx	Primary tumour not assessed	**Tx**	Primary tumour not assessed
T0	No evidence of primary tumour	**T0**	No evidence of primary tumour
Tis	Carcinoma in situ	**Tis**	Carcinoma in situ
T1	Tumour invades submucosa	**T1**	Tumour invades submucosa
T2	Tumour invades muscularis propria	**T2**	Tumour invades muscularis propria
T3	Through invades through the muscularis propria	**T3**	Through invades through the muscularis propria
T4a	Tumour directly invades adjacent organ	**T4a**	Tumour invades visceral peritoneum
T4b	Tumour invades visceral peritoneum	**T4b**	Tumour directly invades adjacent organ
Regional lymph nodes (pN)			
Nx	Lymph nodes cannot be assessed	**Nx**	Lymph nodes cannot be assessed
N0	No lymph node containing tumour cells	**N0**	No lymph node metastases
N1	Tumour cells in 1-3 lymph nodes	**N1a**	Metastases in 1 lymph node
		N1b	Metastases in 2-3 lymph nodes
		N1c	No regional lymph nodes but tumour deposits present in subserosa, mesentery or mesorectum
N2	Tumour cells in 4≤ lymph nodes	**N2a**	Metastases in 4-6 lymph nodes
		N2b	Metastases in 7≤ lymph nodes
Distant Metastases (pM)			
M0	No metastases	**M0**	No distant metastases
M1	Metastases to distant organ	**M1a**	Metastasis confined to one organ or site without peritoneal involvement
		M1b	Metastases in two or more sites without peritoneal metastases
		M1c	Peritoneal surface metastasis

Table 1.1: The 5th and 8th editions of the TNM colorectal cancer classification (14).

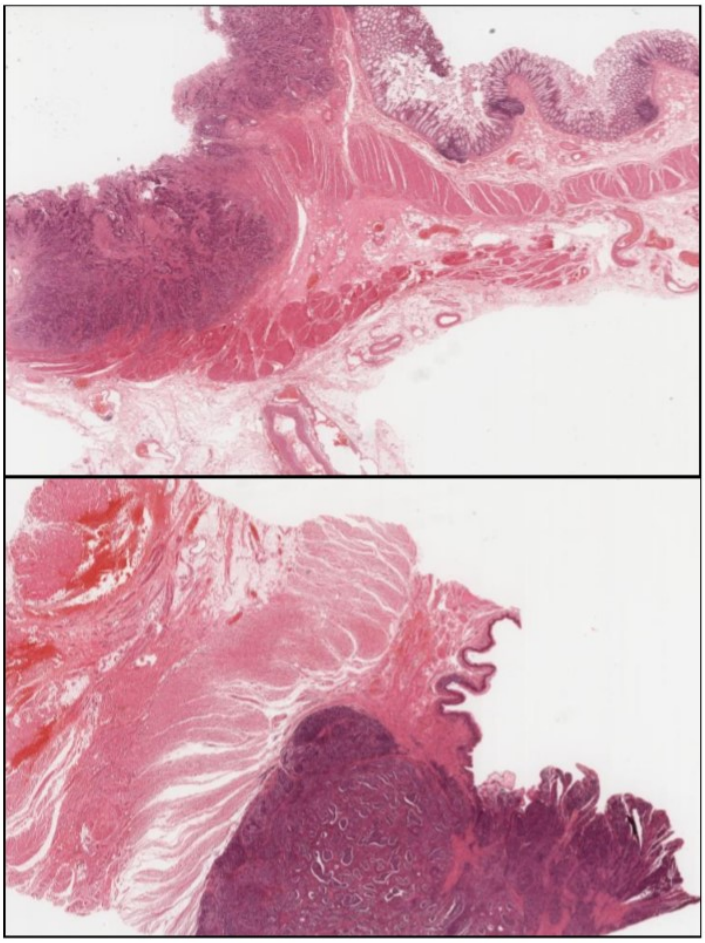


Figure 1.2: Two haematoxylin and eosin stained slides of resected rectal cancers. In the top image, the cancer is seen in the upper left quadrant with evidence of direct muscularis propria invasion. Normal mucosa is seen in the top right. In the lower image, the well-defined cancer edge abuts the muscularis. Images reproduced with written permission from the University of Leeds virtual pathology library (http://www.virtualpathology.leeds.ac.uk/) (appendix 1).

1.3 Surgical resection of the rectum

Whilst multi-disciplinary rectal cancer care continues to evolve, surgical excision remains a mainstay of treatment and offers the best chance of a obtaining a cure. However, primarily due to the presence of too much cancer or patient frailty, only half of UK rectal cancer patients undergo major resection with curative intent (9).

Rectal cancer presents a number of additional challenges not encountered in colonic cancer cases. Anatomically, access to the mesorectum is hampered by a combination of the bony confines of the pelvis both at the pelvic inlet and the lateral side walls. In the low pelvis the distal rectum follows the

sacral curvature. In the supine position and viewed from the abdomen it turns anteriorly and superiorly to pass through the pelvic floor musculature (6). Surgical access is further impaired by the pubic symphysis and the contents of the anterior and central pelvic compartments necessitating approaching the pelvis from the abdomen presenting ergonomic issues for the surgeon in these typically long operations.

Previous problems and the evolution of rectal cancer management

In the 1980 and 1990s poor patient outcomes following rectal cancer surgery were of great concern. Local tumour recurrence rates following resection performed with curative intent was reported between 15-25% even in established centres (18-21). Local recurrence is associated with significant quality of life impairment in addition to the detrimental impact upon overall survival (22-25).

The unacceptable outcomes observed led to an urgent drive to identify the causes and deliver widespread improvements. Multi-disciplinary rectal cancer care has significantly evolved with a number of alterations in core concepts and practices particularly in the pre-operative setting.

Staging

The relatively stable position of the rectum allows for additional staging of the soft tissues of the mesorectum with MRI. Additionally, image degradation due to peristalsis and respiratory artefact during the long image acquisition phases is reduced. In the seminal MERCURY I and II studies, MRI was seen to accurately stage rectal cancer tumour depth, likely nodal involvement and further information on tumour height and quadrantic location (26, 27). Critically, MRI can accurately predict whether the mesorectal fascial plane is free of tumour and therefore whether a complete clear resection is possible (27-29). MRI soft tissue resolution and bowel wall definition greatly exceeds that of pelvic CT images which are now solely performed to assess for the presence of distant metastases and concurrent pathology. MRI mesorectal staging is considered standard of care following endorsement from a number of regulatory and surgical bodies (15, 30, 31).

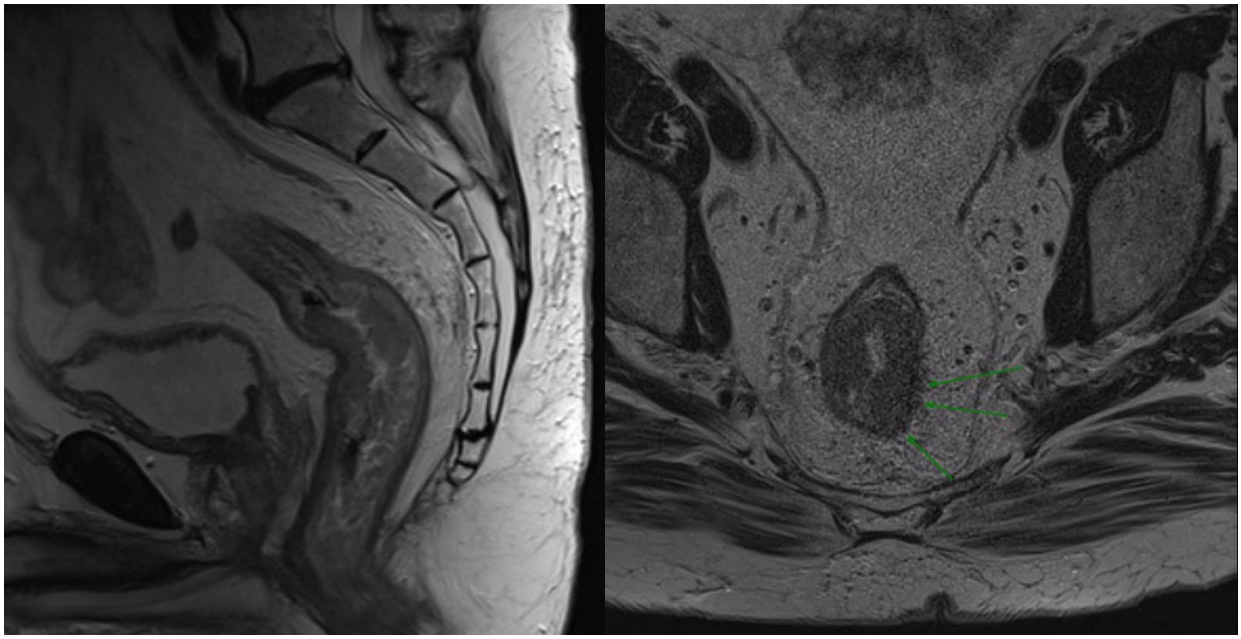


Figure 1.3. T2 weighted MRI images of a mid-rectal adenocarcinoma in a male patient. a) The midline sagittal image shows an annular cancer. b) Axial image perpendicular to mesorectum through the level of the tumour. There is a blurring of the muscularis propria (green arrows) suggestive of T3 disease extension. Histology confirmed pT3b with cancer cells 2mm beyond the muscularis at that point. Original deidentified image kindly supplied by Dr Cen Thomas.

Neoadjuvant treatment

The fixed position of the mesorectum also presents the opportunity to add radiotherapy to the treatment armamentarium. Accurate and repeated tumour targeting is possible allowing administration of therapeutic fractionated doses and importantly the pelvic location reduces the dose delivered to sensitive adjacent tissues particularly the small bowel reducing iatrogenic side effects. Radiotherapy was initially delivered post-operatively to patients assessed as being at a high risk of recurrence. This was superseded by the findings of the German trial where pre-operative neoadjuvant radiotherapy was associated with higher rates of clear surgical margins, reduced toxicity and halved the subsequent lower local recurrence rate albeit with no change in overall patient survival (32). This finding has been reproduced in other major randomised trials with long term follow up data notably CR-07 (33).

There remains debate on the merits of neoadjuvant short course radiotherapy vs. long course neoadjuvant chemoradiotherapy (34). The UK NICE clinical guidelines on the diagnosis and management of colorectal cancer and matched quality standard groups rectal patients into low, moderate and high risk of local recurrence with tailored use of neoadjuvant treatments (15). These guidelines favour the use of synergistic administration of neoadjuvant chemotherapy with fractionated external beam radiotherapy for moderate and high risk cases. Downstaging the cancer aims to allow delivery of a curative resection with clear margins although the optimal timepoint for surgery after completion of neoadjuvant therapy remains debated (35). 38% of rectal cancer patients in England and Wales undergo neoadjuvant treatment followed by major resection although this varied widely across different regions (23-58%) (9).

MDT working practices

Multi-disciplinary team working is mandatory for UK cancer clinicians with patients discussed at various points of their care pathway. MDT working aims to ensure individualised care plans are created using expert consensus based upon best practices and evidence based guidelines. The core members of a colorectal MDT comprise colorectal surgeons, histopathologist, radiologist, oncologist (medical and/or clinical), colorectal cancer nurse specialist and an administrator although modern MDTs typically include representation from extended members including gastroenterology, screening programme representative, research nurse, dietitian, secretarial support, hepatobiliary and thoracic surgeons.

1.4 Surgical Management - Anterior resection of the rectum

The above strategies all play important roles in care improvement but arguably the single largest forward leap came through the improvement in surgical techniques and intervention delivery.

Total mesorectal excision – procedural steps

Heald and colleagues revolutionised the surgical management of rectal cancer (36). In 1986 they reported their prospective observational series of 115 consecutive patients that underwent

complete excision of the rectal mesentery with long term follow up data (37). Termed total mesorectal excision (TME) this technique was based on sound anatomical and surgical oncological principles. They described a sharp circumferential dissection meticulously respecting the enveloping mesorectal fascial plane with adequate proximal and distal mural margins. This aims to ensure all potentially involved lymphatics and tumour nodules are excised and the possibility of shedding viable cancer cells is reduced by respecting the fascial covering layer.

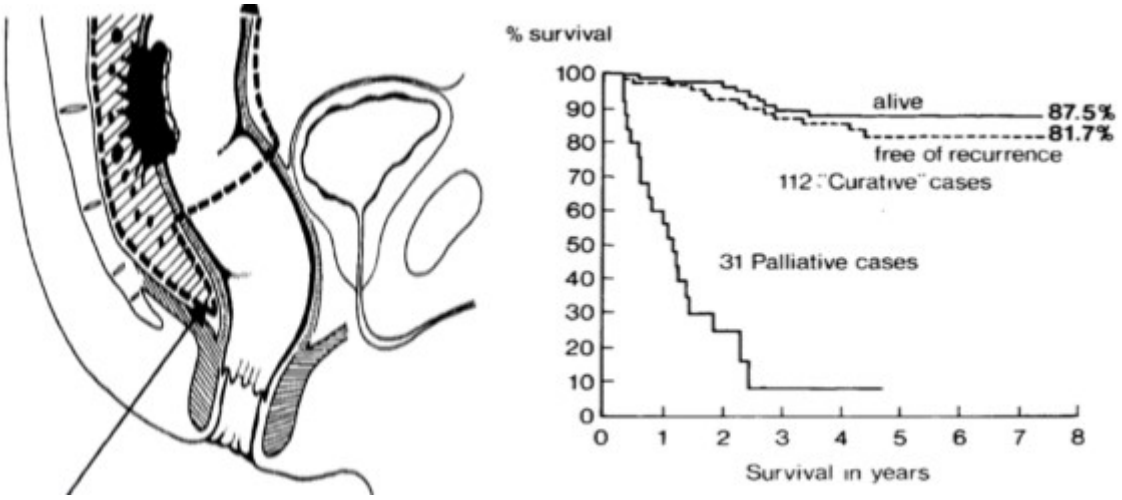


Figure 1.4: Figure depicting the advocated surgical dissection strategy in the original total mesorectal excisions report from Heald et al (36, 37). Long term survival data was far superior to comparable reports at that time. Written permission from reproduction of figures was obtained (appendix 1).

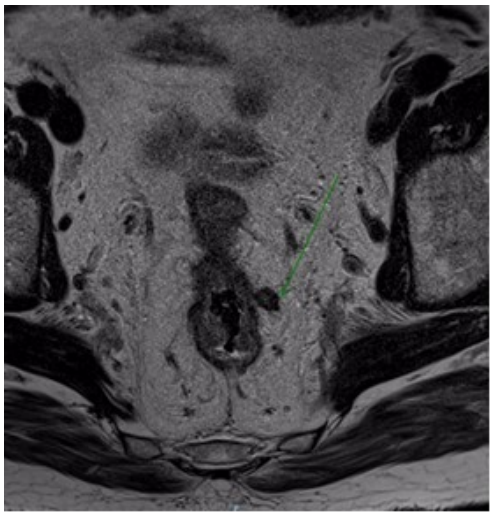


Figure 1.5: T2 weighted axial MRI slice showing direct tumour expansion into the mesorectal fat (arrowed). Intramesorectal surgical excision would risk exposing or leaving viable tumour cells in situ. Meticulous preservation of the enveloping mesorectal fascia reduces this risk.

A number of additional steps are required in TME surgery. High division of the IMA is performed to maximise the number of harvested lymph nodes to both facilitate accurate histopathological staging of the cancer and minimise the chance of leaving active sites of cancer in situ. Secondly, the IMA would be expected to act as a tether preventing translocation of the mobilised proximal colon down to the pelvis to perform a tension free anastomosis. After IMA ligation, the descending colon is fully mobilised from the retroperitoneum along Toldt's fascia and peritoneal attachments to the lateral abdominal wall. When required to gain sufficient mobile length, the splenic flexure of the colon is taken down through division of the greater omentum and disconnection of the mesentery from the inferior border of the pancreatic body and tail. The descending colon maintains its vascularity from the marginal artery of Drummond supplied by the left branches of the middle colic artery that arises from the SMA. An anastomosis is formed at the level of the pelvic floor, typically using circular stapling devices. Surgeons may also create a proximal defunctioning stoma to temporarily divert the faecal stream to assist anastomotic healing and reduce the consequence of a leak. A detailed expert Delphi consensus that aimed to fully standardise the procedural steps of laparoscopic TME surgery has been previously reported (38).

TME and patient outcomes

This theoretical reduction in local and distant recurrence has been proven by numerous clinical studies leading to TME being widely considered as the gold standard surgical procedure for rectal adenocarcinoma. The so called TME era, in combination with the important alterations in multi-disciplinary and peri-operative care, has seen reported local recurrences rates fall to around 4-6% (39-43), dramatically reduced from the peaks of 25% reported in the 1990s (19-21). No randomised trials have been performed on this topic and such a study would now be considered unethical. Nevertheless, there remains a pressing need to further drive down the rate of involved margins following TME resection.

Quality of life and pelvic nerves

The improvement in cancer outcomes and the associated increased numbers of survivors have allowed some attention to move towards improving quality of life following successful treatment. Impaired bladder, bowel and sexual function are common sequalae of uni and multimodality rectal cancer treatment and all can significantly impair quality of life (44-47). An association with injury to the autonomic hypogastric or pelvic splanchnic nerve plexuses that typically overlie the promontory and body of the sacrum respectively in proximity to the posterior mesorectal fascia has been reported. Male urogenital dysfunction is associated with autonomic nerve injury at the level of the prostate where the nerves track anteriorly from the endopelvic fascia (48). Controversy remains regarding both the anatomy of Denonvillier's fascia and the appropriate anterior TME dissection plane (49). Nerve identification and preservation can alleviate much of this risk and therefore forms part of the TME surgical strategy (38, 50).

1.5 Minimal access surgery

Arguably the single largest impact on surgical practice in the last twenty years has come from the widespread uptake of minimal access surgery. Whilst a number of historical devices are described, the modern understanding of minimal access surgery developed rapidly in the late 1980 (51). Production of laparoscopes, containing a light source and optical sensor, allowed digital image processing and monitor projection of detailed images. These were delivered through 10-12mm abdominal ports alleviating the need for large surgical incisions to provide direct line of sight vision. The use of laparoscopic instruments through small ports allowed delivery of abdominal, pelvic and later thoracic surgical interventions without the need for open access.

Laparoscopy developed many early advocates in the view of the reported short-term patient benefits (52, 53). Reduction in tissue trauma and length of incisions resulted in reduced stress response to surgery, analgesic requirements, a more rapid return of function, mobilisation, time to discharge and return to normal activities (54). Arguably the single most important observation

following laparoscopic procedures has been the lower complication rate reported compared to open surgery (52). A widespread and rapid adoption of laparoscopic techniques by gynaecologists, general surgeons and urologists in particular resulted.

The roll out of laparoscopy followed the classical technology adoption life cycle curve (55). The initial excitement was tempered by the realisation that MAS presented new challenges to experienced open surgeons. Adapting to using rigid instruments with reduced haptic feedback combined with the initial limitations of available technology, resulted in the realisation that a proficiency learning curve existed (56). Surgical enthusiasm was curtailed by reports of bile duct injuries during laparoscopic cholecystectomies significantly in excess of those reported in open cases (57). This promoted an awareness and expansion in training initiatives which gradually overcame these issues as surgical experience and understanding increased. In the present day the use of laparoscopy is near total for common abdominal procedures such as cholecystectomy (58). In expert centres, boundaries continue to be pushed with the successful use of laparoscopy in procedures that once appeared only possible as open procedures such as oesophagectomy, thyroidectomy and abdominal aortic aneurysm repair (59-61).

Laparoscopic colon cancer resection

The first laparoscopic segmental colectomy was reported in 1991 but led to intense debate about the use of minimal access surgery for cancer cases (62). A number of concerning reports described inadequate cancer operations and the new finding of port site metastases (63, 64). This discussion continued until the availability of high level evidence provided by a number of major multi-centre, randomised controlled trials that reported equivalence with open surgery. In a subsequent Cochrane systematic review and meta-analysis of 23 randomised trials containing 3526 patients, laparoscopy was associated with less blood loss, post-operative pain, ileus rate and length, pulmonary complications and importantly reduced total and surgical morbidity (65). Improvements in early post-operative pulmonary function and quality of life were also seen after laparoscopic surgery. Long

term follow up data provided by 12 trials (3346 patients) confirmed oncological safety as no difference in local recurrence (colon 5.2% vs. 5.6%, OR 0.84(95%CI 0.47-1.52), p=0.57; rectum 7.2% vs. 7.7%, OR 0.81 (95%CI 0.45-1.43), p=0.46) or cancer specific mortality (colon 14.6% vs. 16.4%, OR 0.8(95%CI 0.61-1.06), p=0.15; rectum 9.2% vs. 10%, OR 0.66(95%CI 0.37-1.19), p=0.16) was seen between the open and laparoscopic cases (65). Minimally invasive surgical approaches for colonic cancer have subsequently been widely utilised due to functional recovery benefits and absence of deleterious effects upon oncological outcomes (53, 65-69).

In 2006 the UK NICE guidelines supported the use of a laparoscopic approach for colon cancer in patients where either approach is appropriate and where sufficient skills are available (70). However, at that time there was insufficient national expertise to allow widespread roll out of laparoscopic techniques. In response, LapCo, a UK specialist training scheme was created and funded by the Department of Health (71). UK colorectal consultants were eligible to undergo supervised in-reach and outreach laparoscopic training with credentialed mentors and underwent defined formative and summative assessment with view to delivering safe independent practice. The successful completion of this initiative has assisted the progressive increase in UK colorectal cancer laparoscopy rates. The UK national bowel cancer audit project reports that 71% of major rectal cancer resections are now performed laparoscopically although there remains variation across the cancer networks and regions (9).

It is noteworthy that LapCo identified the need for bespoke tools for specialist surgeon performance assessment (72-74). As discussed, the additional challenges posed by laparoscopic TME surgery could also be expected to necessitate specialist surgical training to allow safe widespread delivery. No such program has been conceived or implemented in the UK.

1.6 Minimal access surgery for rectal cancer

Following the benefits seen in colonic patients, laparoscopy advocates aspired to expand these to those requiring TME for rectal cancer. The first laparoscopic TME was performed in France in

November 1991 by Leroy and colleagues at the IRCAD centre, Strasbourg (75). The uptake of laparoscopic TME was slow and cautious due to the combination of the specimen quality concerns following open resection and the lack of formal training available for this complex procedure (19, 20, 76, 77).

Technological developments

Ongoing improvements in laparoscopic technology are presenting novel opportunities to rectal cancer surgeons. A variety of trans-anal access ports and specialist insufflators are now commercially available and can facilitate trans-anal minimal access surgical techniques including local resections and trans-anal TME (78). The role of these techniques, particularly in the management of early rectal cancer is an area of active research with a number of clinical trials presently recruiting (79, 80). Novel technologies represent another research avenue that aims to improve the delivery of minimal access TME surgery. Initial favourable reports on robotic platforms for TME surgery were not supported by the major ROLAR randomised trial and the role of robotics remains debated (81).

Definition of laparoscopic TME

For this thesis, laparoscopic TME refers solely to trans-abdominal multiport surgery with rigid instruments and an abdominal specimen extraction site. This aims to provide standardisation and is in keeping with predominant UK surgical practice, previously identified areas of need and the surgical techniques employed by the laparoscopic vs. open TME trials discussed below (38). Specifically, trans-anal, robotic, single port, natural orifice extraction sites or abdominoperineal excision of the rectum procedures are not included in this research except where explicitly stated.

Large randomised trials and current controversies

The role of minimally access surgery in the management of rectal cancer is currently contested. Whilst it is clear that a laparoscopic approach is feasible for performing TME procedures, it remains unanswered if acceptable outcomes from using a minimal access approach can be obtained by all surgeons or is applicable to all rectal cancer cases. The available large multicentre randomised controlled trials on laparoscopic vs. open TME surgery have provided conflicting results. The MRC-

CLASSIC trial, performed in the UK between 1996-2002, contained 381 (48%) rectal cancer patients (82). 34% of their laparoscopic TME cases required conversion to open surgery and a non-significant higher CRM involvement rate was seen in the MAS group (12% vs. 6%, p=0.19) leading the authors to conclude that laparoscopy was not appropriate for rectal cancer (82). The higher CRM rate was not associated with any difference in overall or disease-free survival or local or distant recurrence at close of three-year follow up (83). Therefore, the same research group supported the laparoscopic approach to benefit short term patient outcomes. The COLOR II trial, designed as a non-inferiority study to exclude a 5% difference in three year locoregional recurrence rates, recruited 1103 patients from 30 international centres between 2004 and 2010. Similar short-term patient benefits were seen with no difference in histopathologically assessed outcomes (84). This also resulted in near identical follow up data providing further evidence that laparoscopy was non-inferior (85). The COREAN study randomised 340 Korean patients that had received neoadjuvant chemoradiotherapy (86). Short term benefits were seen with no difference in specimen quality between the arms or long-term survival (87). The American ACOSOG Z6051 trial narrowed recruitment to stage II/III disease that had undergone chemoradiotherapy prior to surgery performed with curative intent (88). The study was designed as a non-inferiority trial with the primary endpoint being an 6% difference in a composite endpoint of complete specimen resections. 35 institutions entered 486 patients between 2008 and 2013. Fewer laparoscopic cases produced complete TME specimens (81.7% vs. 86.9%, difference −5.3%, $p_{\text{non-inferiority}}$=0.41) and higher CRM involvement was seen (12.1% vs. 7.7%, p=0.11), therefore non-inferiority could not be supported. Two year follow up data showed equivalent local recurrence and disease free survival between the trial arms (39).

Published at an identical time, the Australasian ALaCaRT study was co-designed with an identical primary endpoint to the Z6051 trial with an -8% non-inferiority threshold (89). This margin was chosen as the maximally acceptable difference that could be accepted. Surgeons were centrally credentialed prior to enrolling 475 patients from 26 centres between 2010 and 2014. Similar results were seen with a successful resection achieved in 82% of the laparoscopic group vs. 89% from the

open arm (difference −7.0%; $p_{non\text{-}inferiority}$=0.38) with higher laparoscopic CRM involvement (7% vs. 3%, p=0.06) (35). Two year follow up data showed no differences in locoregional recurrence (lap 5.4% vs. 3.1%, difference 95%CI -1.5% - 6.1%, lap HR 1.7 95%CI 0.74-3.9) or disease free survival (80% lap vs. 82%, difference 95%CI -9.3%-5.4%) (90). The authors of both trials concluded that a laparoscopic approach for TME surgery cannot be supported as routine standard of care. Of note, both the ALaCaRT and Z6051 studies incorporated multiple quality assurance mechanisms including intervention standardisation and selected central video and pathology case analysis.

Level one evidence from meta-analyses of randomised controlled trials on the role of minimal access rectal cancer surgery is also conflicting. The 2014 Cochrane database review concluded the laparoscopic approach was equivalent as CRM involvement was equal (40) although a more recent meta-analysis of 14 RCTs containing 4000 patients, which included the ALaCaRT and ACOSOG studies, questioned the oncological safety of laparoscopic TME surgery (91). Although no difference was seen in the involved CRM rate (7.9% lap vs. 6.1% open, RR 1.17 95%CI 0.89-1.53, p=0.26), laparoscopic TME was seen to produce fewer complete resections (13.2% vs. 10.4%, RR 1.31 95%CI 1.05-1.64, p=0.02). A sister meta-analysis study assessing short-term patient outcomes confirmed that some of the benefits observed in colonic cancer patients do translate to rectal resections (92). Significantly less blood loss, shorter incisions, earlier return of bowel function and a 1.7 day (95%CI 2.84- - 0.58, p=0.0003) reduction in length of stay was reported although laparoscopic surgery was seen to take 36 minutes longer. Interestingly, this study showed no difference in morbidity or mortality between the approaches (RR 0.98, 95%CI 0.88-1.09, p=0.75), n=12 RCTs, 3313 patients).

A more recent meta-analysis questioned the methodology of the prior studies (93). Including the same 14 RCTs, as the largest three were designed as non-inferiority studies it was argued that a non-inferiority statistical approach was appropriate. When this was performed, the non-inferiority of laparoscopy was shown for involved CRM (difference 0.79% (90%CI -0.46-2.04%), Δnon-inferiority 2.33%, $p_{non\text{-}inferiority}$=0.026) and non mesorectal fascial plane surgery (1.16% (90%CI -0.27-2.59), Δnon-

inferiority 2.85%, $p_{non\text{-}inferiority}$=0.025). This approach showed a significant improvement in short term patient recovery metrics and time to discharge although apart from a reduction in wound infections (8.2% vs. 3.6%, p=0.02), no significant difference in short term morbidity or mortality was seen. Overall, these results have led to widely varying practice amongst surgeons with no consensus reached. A summary of the major outcomes of the reported trials and subsequent meta-analyses is displayed.

Table 1.2 (next page): Table depicting the reported laparoscopic vs. open TME RCTs and meta-analyses. NR=not reported. ↓= significant decrease and ↑=significant increase with a laparoscopic TME approach. - = no difference. *=ALaCaRT used a combined composite endpoint. Both individual endpoints had a borderline significance of p=0.06.

Trial / Author and reference	Location/ recruiting period	Patients	% APER (lap/open)	Conversion (%)	Operative Time	Blood Loss	Lymph Nodes	Length of Stay	30-day mortality	Morbidity	Anastomotic Leak	CRM <1mm	Incomplete TME
Randomised controlled trials													
CLASICC (82, 83)	UK 1996–2002	381	27 / 30	34	↑	NR	-	↓	-	-	-	-	NR
COLOR II (84, 85)	Europe 2004-10	1103	29 / 23	17	↑	↓	-	↓	-	-	-	-	-
COREAN (86, 87)	Korea 2006–9	340	11 / 14	1	↑	↓	-	-	-	-	-	-	-
Zhou 2004 (94)	China 2001-2	171	0 / 0	NR	-	↓	NR	↓	-	↓	-	NR	NR
Braga 2007 (95)	Italy NR	168	8 / 13	7	↑	↓	-	↓	-	-	-	-	NR
Ng 2008 (96)	Hong Kong 1994-2005	99	100 / 100	10	↑	-	-	-	-	NR	NR	-	NR
Lujan 2009 (97)	Spain 2002-7	204	24 / 21	8	↑	↓	↑	-	-	-	-	-	NR
Liang 2011 (98)	China 2004-8	99	49 / 40	1	↑	NR	-	↓	-	NR	-	NR	NR
Gong 2012 (99)	China, 2008-11	138	34 / 29	3	↑	↓	-	↓	-	-	-	-	NR
Ng 2014 (100)	Hong Kong 2001-7	80	0 / 0	8	↑	↓	-	-	-	↓	-	-	-
ALaCaRT 2015 (89, 90)	Australasia 2010-4	475	8 / 7	9	↑	↓	NR	-	-	-	-	-*	-*
ACOSOG Z6051 2015 (39, 88)	USA 2008-13	486	23 / 24	11	↑	↓	-	-	-	-	-	-	-
Meta-analyses of RCTs													
Vennix 2014 (40)	n/a	3528	NR	14.5	↑	↓	-	↓	-	-	-	-	NR
Martinez Perez 2017 (91)	n/a	4034	NR	13.1	NR	NR	-	NR	NR	NR	NR	-	↑
Martinez-Perez 2017 (92)	n/a	4132	NR	13.8	↑	↓	NR	↓	-	-	-	NR	NR
Acuna 2018 (93)	n/a	4234	NR	8.7	NR	NR	-	↓	-	-	-	-	-

1.7 Surgical performance and patient outcome in rectal cancer surgery

Oncological outcomes are intrinsically linked with surgical quality making optimal intervention delivery of great importance. In their seminal publications, Quirke and colleagues showed that both involved CRM and the histopathologically assessed non mesorectal fascia plane of TME surgery were significant risk factors for medium term local recurrence (19, 20, 76, 77). In a randomised trial of 1156 rectal cancer patients investigating the impact of neoadjuvant and adjuvant radiotherapy on local recurrence, at three years 4% of mesorectal fascial surgery, 7% intramesorectal and 13% of muscularis propria cases had developed local recurrence (p=0.0039) (77). This difference was seen to widen at five year follow up. These results have been reproduced by two larger studies. Data from 1382 patients that underwent TME at 65 Belgian hospitals between 2006-2011 showed non-mesorectal fascial plane had a detrimental impact on recurrence, disease free and overall survival (figure 1.6) (101). A post-hoc analysis of 1236 patients from 88 centres that participated in the German CAO/ARO/AIO-04 RCT reported similar findings with non mesorectal fascial surgery being associated with significantly higher locoregional and distant recurrence and lower disease free and overall survival (41). On multivariate analysis plane of surgery was a significant independent risk factor for local recurrence.

These findings emphasis the need for the delivery of optimal surgery to ensure good patient outcomes and that presently there is evidence that this is goal is not sufficiently being met. Observational data from nearly 17000 United States national cancer database patients operated between 2010 and 2011 is concerning as 17% were seen to have an involved circumferential margin (102).The target for all TME surgeons should be complete excision of the cancer with clear circumferential and longitudinal margins and respecting the mesorectal fascia. Where a minimal access approach is considered, these factors remain paramount and must not be amended to facilitate minimal access approaches. The current evidence highlights the need for surgical improvement if MAS TME is to be broadly applied.

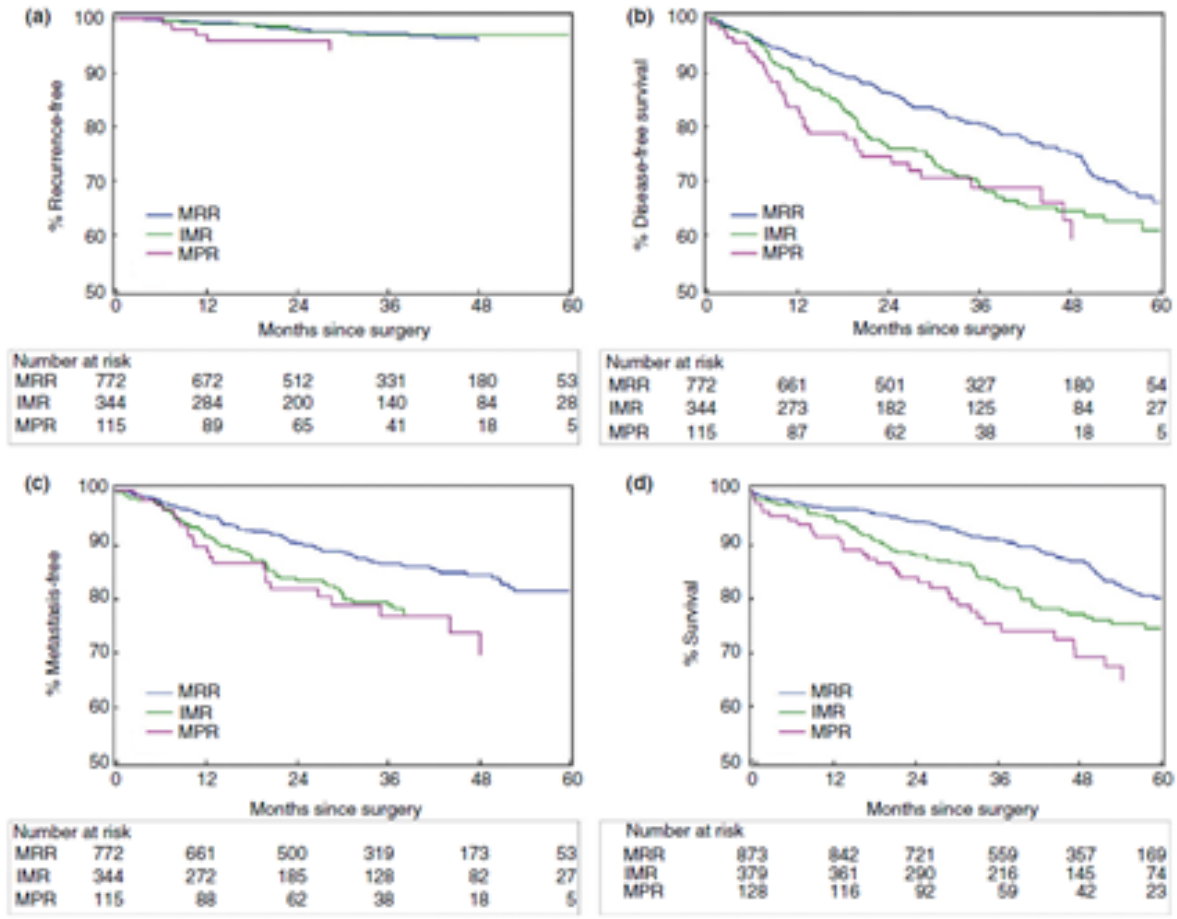


Figure 1.6: Mesorectal fascial (MRR, blue), intramesorectal (IMR, green) and muscularis propria (MPR, purple) Kaplan- Meier survival curves for 1382 patients from 65 Belgian hospitals. (a) local recurrence free survival, (b) distant recurrence free survival, (c) disease free survival and (d) overall survival are all seen to be associated with the obtained plane of surgery. The need for optimal plane, as a surrogate marker for surgical performance, is evident. Figure reproduced from Leonard et al (101) with written permission (appendix 1).

As previous described, the management of rectal cancer presents many challenges and considerations above those seen in colonic cases. Operating on the distal rectum, confined by the bony pelvis and adjacent viscera requires advanced skills, specialist training and completion of a protracted learning curve (56). Long procedures from abdominal access ports present ergonomic difficulties to the surgeon. Maintaining acceptable views can be problematic and clashing of rigid instruments can hamper progress. Tissue changes in response to the cancer or following neoadjuvant chemoradiotherapy use may limit anatomical recognition and further increase the difficulty of the TME dissection.

The need for widespread training was acknowledged and a national MDT training programme for low rectal cancer management "LoREC" was commissioned by the ACPGBI and English national cancer action team (103). Aimed primarily at improving the management of low rectal cancers requiring abdominoperineal excision of the rectum, this programme taught over 1200 clinicians from 90% of UK colorectal MDT's in a series of workshops, MRI and cadaver-based courses (104). The primary goals were to improve MDT decision making and delivery of optimal surgical technique with the view to improving associated patient outcomes including quality of life in addition to oncological data. Comparable national training programmes have been shown to lead to widespread performance improvement with associated beneficial implications for patients (42, 43).

1.8 Assessment theory applied to surgery

As quality of surgery is shown to represent a key outcome determinate and laparoscopic TME trials have reported mixed results, surgical performance assessment could add to our understanding if it exposed reasons underlying the obtained results. Additionally, this could assist efforts to design strategies for improvement and reduce the frequency of suboptimal outcomes. This section explores objective assessment and it use in surgery. Educational theories that underpin the assessment of operative performance are presented.

Assessment comprises of three major educational domains. Cognitive refers to intellectual thought processes, affective covers emotional based principles and psycho-motor skill explores practical task performance (105). Whilst all three domains combine in real world task performance as a craft-based speciality the assessment of psycho-motor physical skills is arguably the most relevant for surgeons (106, 107). As this thesis explores intraoperative surgical performance assessment, focus is deliberately placed upon the psycho-motor domain.

Both formative (assessment during a learning process) and summative (comparing educational outcome and/or performance against a reference standard) assessment are widely adopted in educational practices and directly applicable to the assessment of surgical skill (108). Formative

assessment has been shown to guide and enhance the efficacy of surgical training and reduce the proficiency curve for complex procedures (109). Summative assessment is typically applied as a sign off test to confirm that the appropriate level for safe independent practice has been demonstrated. Competency assessment and credentialing have also been used as a quality assurance technique before contribution to multi-centre surgical trials where variation from differing centres risk introducing potential bias (110). An essential component of summative assessment is the need for an accepted reference standard to be defined to allow performance comparison (105, 108). Establishment and fair comparison with this standard can be a challenging process given the variability in surgical technique, anatomy, pathology and required steps for each individual case. Outside of entry to training interview examinations surgical assessment is entirely criterion-referenced. Demonstration of individual sufficient skill is essential to maintain patient safety. Norm referencing with an arbitrary threshold risks both passing a number of weak candidates and holding back safe trainees who happen to be part of a strong cohort.

1.9 Transition in surgical assessment

For generations, surgical training represented a master-apprenticeship model. Whilst typically based on direct observation, feedback and skill assessment were almost exclusively subjective and opinion based with the training delivered rarely standardised or even reported. Despite its shortcomings this model survived for the majority of 20th century. Criticism came from the lack of objectivity as well as question marks over the validity and reliability of the master-trainee model (111). This outdated system was not in keeping with the increasing scrutiny and regulation of the medical profession.

The craft nature of surgery appears well suited to objective competency assessment where subjectivity is to be avoided. There is an increasing interest from the research and training fields in the objective measurement of surgical performance (112, 113). Comparison against accepted standards and reproducible criteria hold a number of theoretical advantages over the subjective judgement of the trainer, logbook volumes or audited outcomes. Logbook numbers alone give no

indication on the quality of the procedure delivered and subsequent outcome data may be harder to collate and is subject to multiple factors including those entirely out of control of the surgeon. Large numbers, often in excess of the numbers performed in routine practice, maybe required before statistical outliers are identified (114). Understandably, the main focus of assessment has been in the surgical training domain. Formative assessments inform feedback, monitor progress and identify areas of strength and individualised areas to target (105, 115). These methods have been shown to reduce the proficiency gain curve in a number of advanced laparoscopic procedures (109, 116).

Objective assessment methods are structured, unbiased, factual and reproducible measurements based on direct observation with the minimisation of subjectivity and influence in the assessment and grading by the examiner. Objective assessment of operating room performance forms an integral part of both UK and US surgical training programmes (117, 118). Objective techniques, based on the accepted goals for a training programme, reduce bias and improve the reliability of assessments which are likely to be performed by a number of trainers during a rotation (119).

Assessment tools

A number of objective assessment tools have been reported and investigated in the literature including in systematic reviews (112, 113, 120). A mix of generic and task specific scales have been identified with both laboratory (typically box trainer settings) and real world operating room data available. Generic or global scales use domains for performance comparison where each is scored and combined to generate a procedural assessment (121). Checklists, where a task is broken down into technical steps and performance of each is assessed, are a similar assessment strategy for procedural skills such as surgery (122). Tool use aims to remove subjective elements from assessments, allow comparison both of individual performance across time periods and comparison with benchmarking standards and peer group scores as well as assisting trainers by increasing feasibility and usability of objective assessment.

General Surgery (Colorectal) PBA: Rectum-Anterior Resection

APPROVED SURGICAL TEMPLATE Jun 06

Trainee:	Assessor:	Date:
Start time:	End time:	Duration:
Operation more difficult than usual? Yes / No (If yes, state reason)		

The Trainee should explain what he/she intends to do throughout the procedure
The Assessor should provide verbal prompts, if required, and intervene if patient safety is at risk.

Rating: N = Not observed or not appropriate D = Development required
S = Satisfactory standard for CCT (no prompting or intervention required)

	Competencies and Definitions	Rating N/D/S	Comments
I.	**Consent**		
C1	Demonstrates sound knowledge of indications and contraindications including alternatives to surgery		
C2	Demonstrates awareness of sequelae of operative or non operative management		
C3	Demonstrates sound knowledge of complications of surgery		
C4	Explains the procedure to the patient / relatives / carers and checks understanding		
C5	Explains likely outcome and time to recovery and checks understanding		
C6	Includes discussion of potential for stoma formation		
II.	**Pre operative planning**		
PL1	Demonstrates recognition of anatomical and pathological abnormalities (and relevant co-morbidities) and selects appropriate operative strategies / techniques to deal with these		
PL2	Demonstrates ability to make reasoned choice of appropriate equipment, materials or devices (if any) taking into account appropriate investigations e.g. x-rays		
PL3	Checks materials, equipment and device requirements with operating room staff		
PL4	Ensures the operation site is marked where applicable		
PL5	Checks patient records, personally reviews investigations		
PL6	Considers need for temporary/permanent colostomy and confirms site for stoma formation		
III.	**Pre operative preparation**		
PR1	Checks in theatre that consent has been obtained		
PR2	Gives effective briefing to theatre team		
PR3	Ensures proper and safe positioning of the patient on the operating table		
PR4	Demonstrates careful skin preparation		
PR5	Demonstrates careful draping of the patient's operative field		
PR6	Ensures general equipment and materials are deployed safely (e.g. catheter, diathermy)		
PR7	Ensures appropriate drugs administered		
PR8	Arranges for and deploys specialist equipment (e.g. image intensifiers) effectively		
IV.	**Exposure and closure**		
E1	Demonstrates knowledge of optimum skin incision / portal / access		
E2	Achieves an adequate exposure through purposeful dissection in correct tissue planes and identifies all structures correctly		

Figure 1.7: Abridged anterior resection procedure based assessment (PBA) form for UK specialist trainees. Designed for completion by the trainer immediately after observing theatre performance. A large number of generic and specific competencies are graded as satisfactory for completion of training, development needed or not assessable. Reproduced from Beard et al (123) under the UK Government Non-Commercial Licence v2.0.

Assessment of specialists

Public scrutiny of surgeons and their outcomes is increasing. A number of surgical specialities publish individual surgeon and unit outcome data annually as a method of increasing transparency and enhancing public trust. Whilst this principle, in my humble opinion, is to be commended there are a number of factors not explored when mortality data alone is used as seen for colorectal cancer surgeons on the ACPGBI website (124). No information is provided beyond the headline data meaning the reasons behind a death, surgical caseload or patient risk profiles cannot be compared. This is in contrast to the extensive dataset provided by the Great British and Irish society for cardiothoracic surgery which produces surgeon specific patient risk profiles and breakdown of procedures performed over three years together with matched national average data for comparison (125). Importantly, in both these examples, no data on whom surgeons did not operate on (and why) is given meaning the level of risk adverse behaviour cannot confidently be reported. For colorectal surgeons no data on procedures for any other indication than cancer is centrally collected. Therefore, outcome data alone in its present form is unlikely to be sufficient to assess specialist performance.

TME is a complex procedure performed by experienced surgeons. It cannot be assumed that the assessment of specialists can be delivered in a similar matter to trainees. It is likely that specialists would have already been judged competent using the same tools and techniques and given their increased experience may have achieved higher levels of skill (74). This may invalidate those objective tools as a ceiling effect may be encountered where all specialists achieve high marks and it is not possible to discriminate between levels of performance (126). Specialist performance represents the peak of Miller's hierarchical pyramid where the surgeon delivers the procedure whereas assessment tools are typically designed for the "*shows how*" tier (127). It is noteworthy that the available reviews report that only one identified method has been applied to specialist assessment performance (112).

Implications of specialist assessment

Specialist performance assessment carries a number of considerations not seen in trainee or laboratory settings. Should inadequate scores or other areas of concern be seen, the potential risk to patients suggests there is an argument for restriction of practice with potential financial and reputational impact. Institutional healthcare delivery could be affected again with the potential to negatively impact patients. Inaction and treatment delay also risk patient consequences and raises ethical considerations. This is a tricky and poorly defined area in clinical governance. Comparable high-risk industries such as commercial aviation, military leadership and elite level sport have defined mandatory assessment protocols with the number, nature and criteria established. Sanctions for inadequate performance can include immediate loss of practice rights. These assessments carry high stakes for all involved and must be acceptable, valid and reliable to both the profession and the public given the weight attached to their results. These factors may not have been sufficiently explored or applied within modern surgical practice.

Validity and Reliability

Validity and reliability are essential components of any assessment system. Validity purports to the measurements taking place, both that the correct area is being measured and the technique is appropriate (108). Validity has been described as *"the property of being true, correct and in conformity with reality" (128)*. Validity has many sub-divisions used in educational practices, all with translational application to surgical assessment (table 2.1) (112). An ideal tool should be shown to hold a high validity in each domain.

Validity	Definition
Concurrent validity	Correlation with gold standard test measuring same domain
Content validity	Extent to which the intended content is being assessed by the test
Construct validity	Ability for test to differentiate between good and poor performance
Convergent validity	Correlation with another test measuring same domain
Discriminant validity	Discriminates from a different test measuring a different domain
Face validity	Extent to which the test reflects the real world
Predictive validity	Ability of the test to predict future performance

Table 1.3: The different components of validity assessments are shown.

Reliability goes hand in hand with validity. It refers to the consistency of test and reproducibility of outcomes. Like validity, reliability can be subdivided into a number of domains which can be individually measured (112).

Reliability	Definition
Interitem reliability (Internal consistency)	Correlation of different components of a test
Inter rater reliability	Level of agreement between different assessors
Test-retest (intra-rater) reliability	Ability of test to generate similar results when performed at different time points

Table 1.4: Components of reliability assessment

Assessment methods must also be feasible and meet utility considerations otherwise it may not be accepted. Resource cost in both examiner time, training and availability as well as financial

implications must be appropriate to the planned frequency of use and the educational impact expected. Disparity of any aspect could be expected to significantly curtail engagement and tool use.

1.10 Example of specialist assessment in laparoscopic surgery

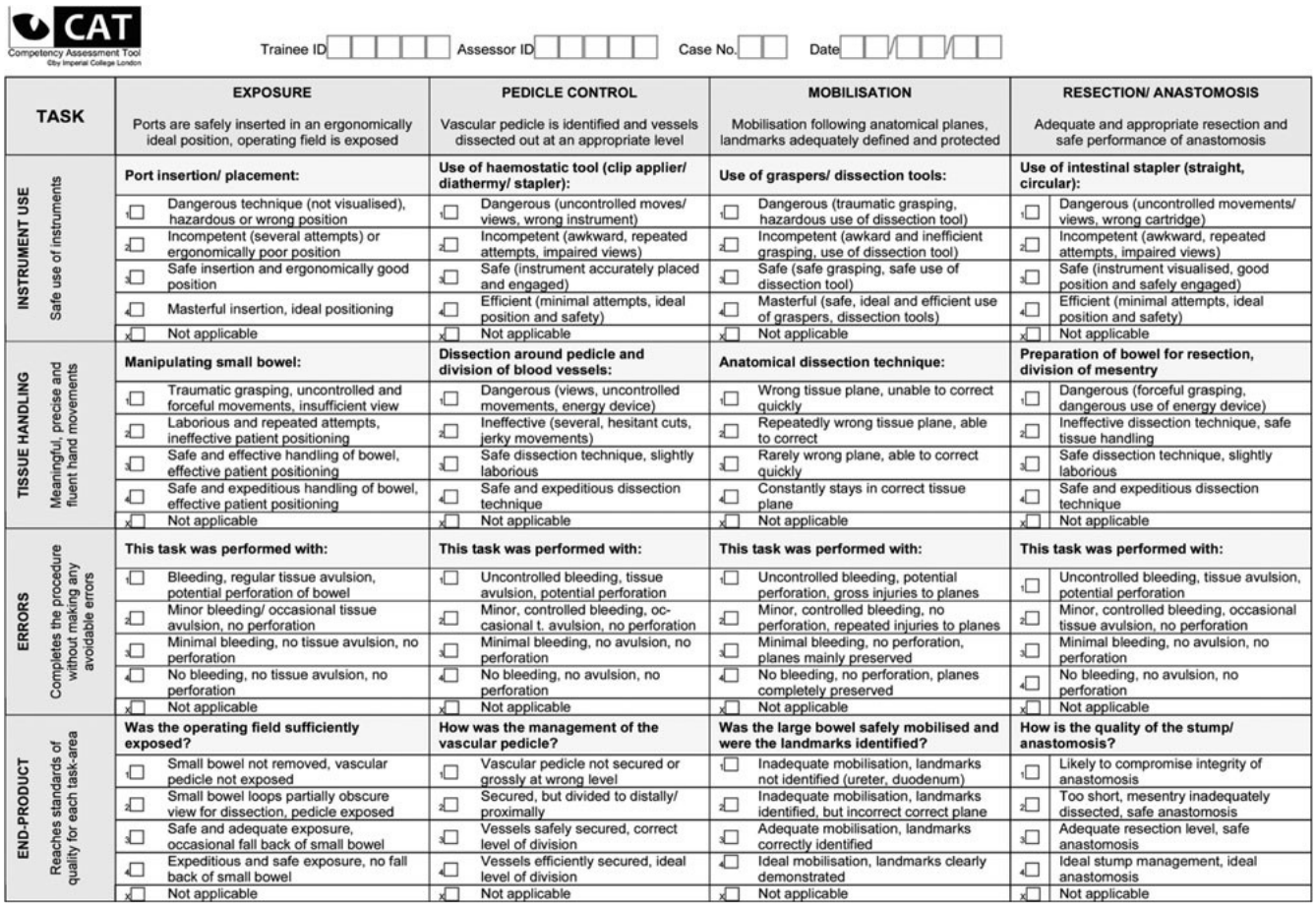
In 2006, the NICE TA105 technology assessment of laparoscopy for colorectal surgery supported use where either an open and laparoscopic approach was appropriate but importantly *"where the surgeon has been trained in laparoscopic surgery for colorectal cancer and performs the operation often enough to keep his or her skills up to date"* (70). It was estimated that only 5% of UK surgeons met this criterion preventing a rapid and safe roll out of minimal access surgery to colorectal patients. In response, the Department of Health funded the LapCo national training programme that aimed to train 250 senior surgeons to enable widespread safe delivery of laparoscopic colorectal surgery in the UK. The specific details of the programme have been reported previously (71).

Of key relevance, assessment of specialist surgeons performing laparoscopic colectomies required the design, validation, pilot and incorporation of two bespoke tools (72, 74). Previously reported instruments were felt not to be applicable to specialist performance as most were designed and had their validity and reliability established on trainees. The possibility of a ceiling effect existed if they were to be applied to specialist performance.

The LapCo steering committee desired a tool to monitor individual progression within the programme. Based on the level of trainer support needed to complete each generic task step in each training case, measurement of individual proficiency curves was possible (72). Quantifying support was felt to appropriately reflect the degree of performance and reduce subjectivity. From 610 analysed assessments, content and contrast validity and inter-rater, test-retest and interitem reliability was established (72).

As no prior competency-based specialist surgical training programme had been performed, there was an additional need for a novel summative sign off assessment. This also had to be designed as

potentially applicable tools were used for junior trainees and unsuitable for high stakes specialist assessment. The laparoscopic competency assessment tool (L-CAT) was created using a Delphi process featuring a number of expert surgeons and met utility requirements (74). Construct validity was established through blinded case video review of expert surgical trainers and participating trainees within the LapCo programme. L-CAT scores differentiated the surgeons including training that passed or failed subjective summative assessment with reliability shown. An average item score of 2.7 was shown to be an appropriate cut off with a 94% sensitivity and 85% specificity with the subjective pass/fail grade awarded by the trainers. Improvements in diagnostic accuracy was seen when the L-CAT was combined with video assessment with 100% sensitivity reached allowing the authors to conclude that the L-CAT allowed summative assessment of specialist surgical performance. Using the 2.7 cut off, clinical validity was seen with lower scores being associated with significantly higher complication rates (18 vs. 6%, p=0.02) and lower lymph node yields (74). Additionally, L-CAT scores were not linked to overall laparoscopic colorectal case experience, but the number of training operations undertaken within the LapCo programme.



CAT
Competency Assessment Tool
©by Imperial College London

Trainee ID ☐☐☐☐☐ Assessor ID ☐☐☐☐☐ Case No. ☐☐ Date ☐☐/☐☐/☐☐

TASK	EXPOSURE Ports are safely inserted in an ergonomically ideal position, operating field is exposed	PEDICLE CONTROL Vascular pedicle is identified and vessels dissected out at an appropriate level	MOBILISATION Mobilisation following anatomical planes, landmarks adequately defined and protected	RESECTION/ ANASTOMOSIS Adequate and appropriate resection and safe performance of anastomosis
INSTRUMENT USE Safe use of instruments	**Port insertion/ placement:**	**Use of haemostatic tool (clip applier/ diathermy/ stapler):**	**Use of graspers/ dissection tools:**	**Use of intestinal stapler (straight, circular):**
	1 ☐ Dangerous technique (not visualised), hazardous or wrong position	1 ☐ Dangerous (uncontrolled moves/ views, wrong instrument)	1 ☐ Dangerous (traumatic grasping, hazardous use of dissection tool)	1 ☐ Dangerous (uncontrolled movements/ views, wrong cartridge)
	2 ☐ Incompetent (several attempts) or ergonomically poor position	2 ☐ Incompetent (awkward, repeated attempts, impaired views)	2 ☐ Incompetent (awkard and inefficient grasping, use of dissection tool)	2 ☐ Incompetent (awkward, repeated attempts, impaired views)
	3 ☐ Safe insertion and ergonomically good position	3 ☐ Safe (instrument accurately placed and engaged)	3 ☐ Safe (safe grasping, safe use of dissection tool)	3 ☐ Safe (instrument visualised, good position and safely engaged)
	4 ☐ Masterful insertion, ideal positioning	4 ☐ Efficient (minimal attempts, ideal position and safety)	4 ☐ Masterful (safe, ideal and efficient use of graspers, dissection tools)	4 ☐ Efficient (minimal attempts, ideal position and safety)
	x ☐ Not applicable	x ☐ Not applicable	x ☐ Not applicable	x ☐ Not applicable
TISSUE HANDLING Meaningful, precise and fluent hand movements	**Manipulating small bowel:**	**Dissection around pedicle and division of blood vessels:**	**Anatomical dissection technique:**	**Preparation of bowel for resection, division of mesentry**
	1 ☐ Traumatic grasping, uncontrolled and forceful movements, insufficient view	1 ☐ Dangerous (views, uncontrolled movements, energy device)	1 ☐ Wrong tissue plane, unable to correct quickly	1 ☐ Dangerous (forceful grasping, dangerous use of energy device)
	2 ☐ Laborious and repeated attempts, ineffective patient positioning	2 ☐ Ineffective (several, hesitant cuts, jerky movements)	2 ☐ Repeatedly wrong tissue plane, able to correct	2 ☐ Ineffective dissection technique, safe tissue handling
	3 ☐ Safe and effective handling of bowel, effective patient positioning	3 ☐ Safe dissection technique, slightly laborious	3 ☐ Rarely wrong plane, able to correct quickly	3 ☐ Safe dissection technique, slightly laborious
	4 ☐ Safe and expeditious handling of bowel, effective patient positioning	4 ☐ Safe and expeditious dissection technique	4 ☐ Constantly stays in correct tissue plane	4 ☐ Safe and expeditious dissection technique
	x ☐ Not applicable	x ☐ Not applicable	x ☐ Not applicable	x ☐ Not applicable
ERRORS Completes the procedure without making any avoidable errors	**This task was performed with:**	**This task was performed with:**	**This task was performed with:**	**This task was performed with:**
	1 ☐ Bleeding, regular tissue avulsion, potential perforation of bowel	1 ☐ Uncontrolled bleeding, tissue avulsion, potential perforation	1 ☐ Uncontrolled bleeding, potential perforation, gross injuries to planes	1 ☐ Uncontrolled bleeding, tissue avulsion, potential perforation
	2 ☐ Minor bleeding/ occasional tissue avulsion, no perforation	2 ☐ Minor, controlled bleeding, occasional t. avulsion, no perforation	2 ☐ Minor, controlled bleeding, no perforation, repeated injuries to planes	2 ☐ Minor, controlled bleeding, occasional tissue avulsion, no perforation
	3 ☐ Minimal bleeding, no tissue avulsion, no perforation	3 ☐ Minimal bleeding, no avulsion, no perforation	3 ☐ Minimal bleeding, no perforation, planes mainly preserved	3 ☐ Minimal bleeding, no avulsion, no perforation
	4 ☐ No bleeding, no tissue avulsion, no perforation	4 ☐ No bleeding, no avulsion, no perforation	4 ☐ No bleeding, no perforation, planes completely preserved	4 ☐ No bleeding, no avulsion, no perforation
	x ☐ Not applicable	x ☐ Not applicable	x ☐ Not applicable	x ☐ Not applicable
END-PRODUCT Reaches standards of quality for each task-area	**Was the operating field sufficiently exposed?**	**How was the management of the vascular pedicle?**	**Was the large bowel safely mobilised and were the landmarks identified?**	**How is the quality of the stump/ anastomosis?**
	1 ☐ Small bowel not removed, vascular pedicle not exposed	1 ☐ Vascular pedicle not secured or grossly at wrong level	1 ☐ Inadequate mobilisation, landmarks not identified (ureter, duodenum)	1 ☐ Likely to compromise integrity of anastomosis
	2 ☐ Small bowel loops partially obscure view for dissection, pedicle exposed	2 ☐ Secured, but divided to distally/ proximally	2 ☐ Inadequate mobilisation, landmarks identified, but incorrect correct plane	2 ☐ Too short, mesentry inadequately dissected, safe anastomosis
	3 ☐ Safe and adequate exposure, occasional fall back of small bowel	3 ☐ Vessels safely secured, correct level of division	3 ☐ Adequate mobilisation, landmarks correctly identified	3 ☐ Adequate resection level, safe anastomosis
	4 ☐ Expeditious and safe exposure, no fall back of small bowel	4 ☐ Vessels efficiently secured, ideal level of division	4 ☐ Ideal mobilisation, landmarks clearly demonstrated	4 ☐ Ideal stump management, ideal anastomosis
	x ☐ Not applicable	x ☐ Not applicable	x ☐ Not applicable	x ☐ Not applicable

Figure 1.8: The Laparoscopic colectomy competency assessment tool. Bespoke design to meet the requirements of the UK LapCo programme that required reliable and valid assessment of complex procedures performed by specialists. Figure reproduced from Miskovic et al (74) with written permission (appendix 1).

The LapCo experience showed assessment of complex laparoscopic procedures at the specialist level is possible and was seen to be both a valid and reliable technique. It is possible that these findings are directly applicable to the use of laparoscopy in rectal cancer surgery, but the L-CAT and GAS tools were not designed or used for this purpose and their content does not cover TME specific considerations. Task specific instruments appear indicated given the different anatomical and procedural steps of TME surgery. In his thesis, Mr Jake Foster performed an initial exploratory investigation into assessment of specialist TME performance and concluded that *"objective assessment of laparoscopic rectal cancer resection is feasible but complex"* (129).

1.11 Assessing surgery

Surgical skill is of paramount importance as performance can directly impact patient outcomes (130, 131). Traditionally, surgical outcomes are reported using post-operative metrics. Routinely collected data such as histopathological, morbidity and mortality measures are widely used as surrogate markers of surgical quality with some UK results published in the public domain (9, 124). However, these can be indiscriminate and indirect variables which may not be reliable or appropriate metrics for direct assessment of surgical skill. Morbidity and mortality can be the result of multi-factorial events that may involve other multi-disciplinary clinicians and factors entirely outside of the operating theatre or the surgeon's control. These methods also ignore non-consequential intraoperative surgical errors and little information on causative events or identification of areas for development provided (132). It could therefore be argued that judging an operation solely outside of theatre is insufficient particularly in high stakes settings. Consequentialism and acceptable ends (such as obtaining a good TME specimen and acceptable short term patient outcome) should not alleviate consideration of the means behind these outcomes. Post-operative data such as histology is only known and hence applied retrospectively on an individual patient basis with no proven predictive validity for future procedures. Scrutiny of the intraoperative period could help quality assure the delivery of surgical interventions including MAS and TME procedures where a need for improvement is evident. This may present secondary benefits to future patients through identification and intervention of areas of deficiency. Arguably post-operative assessment alone cannot sufficiently achieve this goal.

The assessment of surgical performance is common and most frequently applied in the residency training setting. Documented formative assessments are an integral part of UK and US surgical training programmes (118, 133, 134). A number of procedure specific objective tools have been described for this purpose (107, 135, 136). Given the challenges and current literature outlined,

laparoscopic TME can be expected to be performed by specialists rather than trainees so when assessment is planned, it is essential the methods used are valid for this level (74, 126).

Objective evaluation of intraoperative performance through observation of the surgeon represents an underused method of assessing the true delivery of each intervention. Laparoscopic platforms facilitate intraoperative performance assessment with advantages not available in open surgery. Utilising the integrated digital image capture systems incorporated into most systems presents new opportunities for objective assessment through direct and reproducible observation of performance as well as facilitating unbiased, blinded and repeated review (137, 138). The maximal face validity also represents the optimal setting for the assessment of surgical performance with relevance to formative training and summative surgeon credentialing. Within a research setting, video review can aid quality assurance and standardisation of intervention delivery, for example within multicentre studies and is likely to represent the most important assessment time point for investigating new surgical technologies or procedures (110, 139). Such a strategy may also reveal any association between the intervention and subsequent observed outcomes. Although rarely used to date video based assessments have been successfully employed as primary endpoints in rectal cancer trials (140) and training programmes (73, 74, 132).

Compared with post-operative assessments, intraoperative review is in its infancy. Subjective assessment by a trainer was commonplace and formed the backbone of 20th century surgical training. However, such a strategy is open to significant criticism and is now widely considered outdated and unfit for purpose as the primary measure of proficiency (111). Objective assessment is desirable to reduce subjectivity, facilitate peer comparison for both the trainee and assessor and ensure required standards are being met (113). Use of intraoperative assessment tools has increased and been shown to be valid and reliable for a number of procedures and different levels of surgeons including specialists performing laparoscopic colectomies (72-74, 112, 131). The UK LapCo

programme focussed on the delivery of colonic resection training meaning no such instruments may exist for laparoscopic TME surgery.

1.12 Quality assurance of operative performance

Quality assurance is defined by the Oxford English dictionary as “the maintenance of a desired level of quality in a service or product, especially by means of attention to every stage of the process of delivery or production”. QA is a core concept in many industrial and production-based systems. It represents a systematic process investigating whether the service meets defined quality criteria. The fundamental QA steps and the requirement to meet them have been defined by the International Organisation for Standardisation guidelines (ISO 9000 and 9001). These represent the most widely adhered to management tools with over one million organisations achieving certification in a number of industry specific adaptations including the manufacture of medical devices.

Quality control complements quality assurance through its focus upon the factors involved in the service, production or intervention delivery. Ensuring appropriate human factors such as skills sets, teamwork and effective management are in place is a key requirement for complex systems particularly those based upon human interaction including healthcare delivery and individual operative procedures. Quality control aims to ensure that the service delivered meets the needs of the patient as well as the defined quality standards for that specific task.

On face value, these concepts appear directly translatable to healthcare delivery but an overview of quality assurance in the NHS provided by the independent Kings Fund charity focussed almost exclusively on organisational and regulator level factors (141). Staff morale and engagement were acknowledged as critical steps, but the document can be argued to be far removed from routine patient care and the operating theatre.

A more focussed exploration on surgical interventions is required to assess the impact of surgical skill and intervention delivery and clinical outcomes. The limited number of operating theatre

reports come primarily from upper gastrointestinal procedures. The Michigan Bariatric Surgical Collaborative studied single operative videos from 20 surgeons that were reviewed by ten peers. A wide variation in technical skill was reported with strong clinical validity shown for operative time, short term morbidity, re-operation, readmission and mortality (130). A retrospective video review of 61 laparoscopic gastrectomy procedures showed technical performance was an independent predictor of major short term patient outcomes (142). In a systematic review of 7775 procedures within 24 studies, 21 studies concluded superior technical performance positively affects short and long term morbidity and mortality (143). It is therefore surprising that surgical quality assurance and patient safety focuses on systems and their contributions over individual technical skill. As a principle it appears logical to consider the surgeon as the driver of outcomes through their technical performance and therefore warrants a higher profile in the quality assurance and research fields.

1.13 Application of quality assurance and quality control to colorectal surgery

The definition of quality assurance can equally be applied to surgical interventions. Unlike industrial processes where a high degree of automation or standardisation is typically commonplace for example in manufacturing production lines there are a large number of non-comparable factors that may lead to widespread variation in surgical interventions. The need for standardisation is highlighted in the CONSORT checklist extension for non-pharmaceutical interventional trials although no suggestions on how to reduce these bias risks are offered (144). Different quality assurance mechanisms are likely to be required for surgical assessment. Surgical procedures are frequently personalised in response to the pertinent factors in each case particularly anatomical ones. In rectal cancer, the tumour height alone represents a significance source of variation as could pelvic dimensions which also vary between patients and genders (1, 145).

In their guidelines for the management of colorectal cancer, a number of bodies provide standards, quality indicators and practice guidelines (146). In 2012 the UK National Institute for Clinical Excellence produced quality standards for colorectal cancer care (30). The main focus of these

documents was quality control and system-based processes. Surgery is assessed solely outside of the operating theatre. The document makes no mention of operative procedures and or the quality of delivery despite the proven links with oncological outcomes. It is notable that just two of the 21 strong NICE working group were colorectal surgeons. Specifically there is no quality standard for colorectal surgical performance or assessment (30). This runs against evidence of the direct links between specimen quality and outcomes and its surprising that this routinely reported, and easily collectable data is not subject to regulatory assessment or targeted improvement (77, 101).

Similarly, in the 2017 ACPGBI colorectal cancer guidelines the quality of surgery receives no attention (147). Relevant recommendations are limited to *"surgery for colorectal cancer should be performed by surgeons with appropriate training and experience, working within an MDT"* and *"choice of rectal resection should be tailored to the individual patient, focussing on achieving complete resection, low morbidity and restorative procedures in appropriate cases."* The sole recommendation for outcome reporting and audit is *"surgeons and trusts must make provisions for the prospective collection of accurate clinical data for submission to the National Bowel Cancer Audit (NBOCA)."* Their core dataset omits any review of the intraoperative period or linked factors (148).

A systematic review identified wide heterogeneity in reported colorectal quality indicators with the majority being systems based processes (149). As the focus on surgical quality control appears to lie away from the technical performance and intervention delivery it is perhaps unsurprising that the results of such initiatives have had a limited impact on clinical outcomes (138, 150). Operative performance assessment appears appropriate as video performance assessment within the LapCo programme showed significant associations with both patient and histological outcomes (131).

This situation appears different within the research setting. A number of mechanisms were employed by completed multi-centre laparoscopic colorectal cancer randomised controlled trials (110). Despite the lack of accepted quality assurance methodology frameworks, all were seen to attempt to reduce the risk of performance bias. The review made recommendations:

Mechanism to reduce performance bias	Relevant CONSORT Checklist Items	Examples Identified	Recommendations
Credentialing of surgeons and centres for participation in the study.	3) *"Where applicable, eligibility criteria for centres and those performing the intervention"* 15) *"Where applicable, a description if care providers (case volume, qualification, expertise etc.) and centres (volume) in each group."* 20) *"take into account … unequal expertise of care provides or centres in each group"*	Credentialing based on minimum experience in laparoscopic colorectal resections Description of baseline experience of surgeons Credentialing of expertise based on submission of video recording of laparoscopic operations assessed by expert against protocol	Surgery should only be performed by credentialed surgeons / at credentialed units Eligibility criteria for participation should be agreed by the trial steering committee and detailed in the trial protocol. Credentialing of surgeons / surgical units should incorporate assessment of both experience and expertise in the intervention being studied. For credentialing based upon experience, this should be evidence-based, e.g. using statistical analysis of learning curves to identify thresholds at which proficiency is achieved for relevant outcome measures For credentialing based upon expertise assessment, validated assessment tools should be used with agreed thresholds for eligibility. Where no such tool exists, either a generic tool, or structured checklist could be used with set criteria to be assessed in order to facilitate objective, reliable evaluation.
Standardisation of trial intervention and comparator in the trial protocol	4) *"Precise details of both the experimental treatment and comparator"* 4A) *"Description of the different components of the interventions and, when applicable, descriptions of the procedure for tailoring the interventions to individual participants"* 4B) *"Details of how the interventions were standardised"*	Standardization of technical and oncological aspects of laparoscopic colorectal resection Description of components of laparoscopic procedure in protocols Use of videos and live demonstrations to disseminate techniques	An agreed technique for the surgical procedure being studied should be defined clearly in the protocol by the trial steering committee. The protocol should define the technical steps of the procedure and acceptable variations in technique. Where international consensus has been established on a recommended technique for performance of a procedure, the researchers are strongly encouraged to follow this technique.

			Researchers are encouraged to facilitate dissemination of their protocol technique through workshops and video presentations.
Monitoring of technical performance	4C) *"Details of how adherence of care providers with the protocol was assessed or enhanced"* New Item) *"Details of the experimental treatment and comparator as they were implemented"*	Random audit of video recordings of operations performed within trial Centralized review of pathology with agreed interventions should be defined in the protocol Random sampling of cases should be used and audited for evaluation of technical performance for fidelity against the standardised description of the procedure in the protocol (using validated assessment tools / checklists where applicable) Where clinical endpoints are known to be closely linked to technical performance or failure, an auditing strategy should be described in the protocol with thresholds at which either a practitioner is brought to the attention of the steering committee or stopped from participating in the trial threshold for alerting monitoring committee Monitoring of rates of conversion from laparoscopic to open surgery with agreed threshold for alerting monitoring committee.	A strategy for monitoring technical performance of delivery of the trial

Table 1.5: The mechanisms, examples and recommendations reported to reduce performance bias within colorectal multicentre RCTs. It is apparent that reporting frameworks such as CONSORT are not applicable to surgical intervention trials. Reproduced from Foster et al (110) with written permission (appendix 1).

Through review of published papers and trial protocols, one of the consistently identified themes was the monitoring of surgeons during the trial. The COST and CLASSIC trials required all laparoscopic cases to be recorded unedited and submitted to the central trial office (82, 151).

Randomly selected cases underwent audit to “evaluate sound oncological technique” in keeping with the principle of acceptance sampling used by production industries. If the assessed cases are satisfactory it is assumed the entire series is similarly acceptable. The assessment method was not stated with no mention of objective, reproducible or previously reported techniques. Other methods were centralised review of histopathological data with independent monitoring of margins by data committees. These strategies represent a statistical quality control, where performance within the defined margin is acceptable. There is a potential risk that infrequent adverse outcomes and important information that may describe areas for improvement and future avoidance is missed if only statistical outliers are studied. Reliance on statistics can impact individual patients as their own experience could be defined as 0 vs. 100% for binary endpoints. Overall, QA and QC are applicable to surgery but appear in their infancy with particular scope to expand into the intraoperative period.

1.14 Discussion

This chapter has discussed rectal cancer and the central role of TME surgery in contemporary management. Surgeons have sought to replicate the short term benefits observed following laparoscopic colonic surgery where no negative oncological impact was seen, to patients with rectal cancer. Total mesorectal excision presents many challenges for the surgeon that are not encountered in colonic cancer cases. Presently, the literature questions the role of minimal access surgery for rectal cancer resection but over 60% of UK rectal cancer cases are attempted laparoscopically which highlights the variations in practice that have resulted. Although laparoscopic advocates continue to perform laparoscopic rectal cancer resections with acceptable audited results, the generalisability and wide acceptability of the technique has not been established. Whilst the large randomised trials have shown laparoscopic rectal cancer surgery can be safely delivered in the majority of cases, there remains a poorly defined subset where suboptimal outcomes occur. Patient outcomes are directly linked to the quality of surgery through the completeness of the resected mesorectum making optimal intervention delivery a critical requirement for surgeons.

Assessment of intraoperative performance could assist our understanding of trial results through the measurement of performance and its associated impact upon outcomes. Assessment theory and its application to surgery was explored. The continuing transition away from traditional subjective techniques holds many advantages and helps ensure demonstration of competency with the goal of ensuring patient safety through delivery of high quality interventions.

Although focusing on performance seems logical and intuitive there is very little research on this topic as the measurement of human performance particularly in complex systems and tasks such as surgery can be challenging. Objective assessment of intraoperative performance is likely to require task specific tools. Without structure, assessments risk being subjective, non-reproducible and of limited use which is particularly inappropriate for practicing surgeons undergoing high stakes assessment with associated issues including possible medicolegal or practice implications. Given the difficulties outlined, MAS TME operations are likely to be performed by experienced surgeons meaning assessment instruments should be appropriate and validated for the specialist level.

The present application of quality assurance to specialist performed surgical procedures appears limited to research study design and post-operative factors. This could be considered to fail its intended goal as being both not systematic nor fully investigating surgery if surgical performance and intervention delivery is overlooked. Industrial quality assurance methods do not appear suited to personalised health care delivery as acceptance sampling does not meet the requirements for the application of quality assurance to rectal cancer surgical performance. Sampling may ensure that overall quality is satisfactory but risks missing individual cases where this is not true. It may take a large number of inadequate outcomes, with their associated impact upon patients, before a problem is identified and many more may be needed before causation can be confidently reported. Secondly, the use of statistical process control where a threshold is established such as peer comparison or a dichotomous measure such as lymph node yield above or below 12 is used, carries the same concerns that individual poor cases are accepted if the overall series meets the chosen criteria. As

the aforementioned rectal cancer trials have shown whether open or laparoscopic approaches were performed, the majority of cases were performed adequately, the described features suggest that intraoperative surgical quality assurance should be applied to all cases to maximise both our understanding and the potential improvements.

Overall, quality assurance of the intraoperative period in surgery is underused and under investigated. There is a pressing clinical need highlighted by the presented specimen concerns reported by the TME literature. Presently, there is very little information as to why these results have been obtained. Exploration of the intraoperative period through the dedicated application of quality assurance to laparoscopic total mesorectal surgery is indicated and could represent a new avenue towards quality improvement.

1.15 Conclusions and rationale for this thesis

Minimal access surgery is widely used in the management of rectal cancer but currently disputed by the differing results of major randomised trials. Despite its inherent variability and link with patient outcomes, the intraoperative period particularly surgical performance and intervention delivery receives little scrutiny. Examining surgical performance is required in order to attempt to improve our understanding of outcome differences seen across the reported MAS rectal cancer studies.

The concept of quality assurance for surgical interventions is sound but presently insufficiently described particularly as intraoperative QA has only rarely been performed. Application of quality assurance to the intraoperative period presents a number of potential uses including formative and summative assessment, credentialing before independent practice or participation in surgical trials and interpretation of histopathological and patient outcomes. Importantly, review of this time period may assist interpretation of trial outcomes and inform strategies to overcome identified areas of deficiency. This attractive strategy could assist the optimal delivery of laparoscopic rectal cancer surgery with associated direct benefits to patients.

Therefore, this thesis aims to explore the development, application and evaluation of intraoperative quality assurance of laparoscopic TME surgery through objective performance assessment and evaluation of any associated clinical impacts.

1.16 Hypotheses and objectives of this thesis

The overarching aim of this thesis is the promotion of the oncological safety of laparoscopic TME surgery through the application of quality assurance to the delivery of interventions.

Hypotheses

1) Quality assurance tools for laparoscopic total mesorectal excision can be developed and applied to research and clinical settings.
2) Quality assurance tools can reliably assess operative technique and correlate with clinical outcomes following laparoscopic total mesorectal excision.

Objectives

1) Identify the reported tools for objective assessment of intraoperative laparoscopic TME performance.
2) Develop and apply objective assessment tools to laparoscopic TME randomised controlled trials in order to evaluate:
    i. Tool reliability
    ii. Concurrent validity
    iii. Construct validity
    iv. Clinical outcome validity

Chapter 2 – Evidence synthesis on the objective assessment of laparoscopic total mesorectal surgery

As there is an intrinsic link between the delivery of surgery and patient outcomes, quality assurance of surgical performance could act as an important safeguard. Before the application of intraoperative quality assurance to laparoscopic TME surgery can take place, an exploration of the literature is required with view to identifying and selecting methods for continuing use in this thesis.

This chapter includes evidence synthesis from a systematic review undertaken to identify and critique the reported tools used for intraoperative surgical performance assessment of laparoscopic TME. The most suitable are selected for ongoing use and application within this thesis.

I gratefully acknowledge the contribution of Mr Joseph Davids (University Hospital of Wales, Cardiff) who independently searched the literature, screened articles and extracted the data alongside me and assisted the preparation of the published manuscript.

As described in section 2.6, the development of the laparoscopic TME performance tool was performed by Mr Jake Foster and Mr Danilo Miskovic during their research time in the Department of Surgery and Cancer, Imperial College London with the full reports provided in their respective theses (129, 152).

2.1 Evidence Synthesis - Introduction

In order to assess the intraoperative performance of specialist surgeons during laparoscopic total mesorectal excision, it is reasonable to assume that task specific tools appropriate for this level will be required. Tool development can represent a significant undertaking as shown by the design of the bespoke LapCo L-CAT (74). This was a time and labour-intensive process for both the surgical education researchers and the specialist surgical trainers that designed and performed the operative video assessments using the new tool. Adopting an appropriate existing system could alleviate this step meaning, if available, using previously reported tools represents an attractive strategy but it is unclear if suitable instruments already exist for this purpose. Source data from previous use would aid the establishment of reference standards and allow comparison of performance with an external group of surgeons and their outcomes. If not previously researched, a formal exploration of reliability and validity may be possible within this thesis.

Aims

As it is unknown if any such objective assessment tool designed for the specialist level has been described for total mesorectal excision performed laparoscopically, the aim of this chapter was to systematically review the literature to report the tools used for intraoperative TME assessment. Additionally, to assess their usage and applicability to specialist performed procedures and given the overarching aims of this thesis, assess any previous tool use within the previously reported laparoscopic TME multi-centre randomised controlled trials.

2.2 Methods

A systematic search of the literature was conducted in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) guidelines (153). The PubMed and Cochrane Library electronic databases were searched covering the period between 1 January 1982 and 1 November 2016. The start date of the review period was chosen to coincide with the earliest

reports of TME in the literature. The EMBASE library was not searched as conference abstracts would not contain sufficient tool detail to allow critique or subsequent use.

The search was performed independently by two researchers capturing terms for minimally invasive including laparoscopic, robotic, trans-anal TME (TaTME) and single-port surgery, TME and assessment of performance (table 2.1). Search terms within each set were combined using the Boolean operator “OR” and the three sets combined with the Boolean operator “AND.” After screening titles and abstracts, articles that fulfilled the eligibility criteria were identified and their full-text publications were independently reviewed. Search results were supplemented by a hand search of selected reviews and the reference lists of all included studies to identify further potentially eligible studies.

Minimally invasive laparoscopic.tw OR laparoscope.tw OR laparoscopy.tw OR keyhole.tw OR minimally invasive.tw OR minimally-invasive.tw OR robotic.tw OR robot.tw OR robot assisted.tw OR stereoscopic.tw OR SILS.tw OR single incision.tw OR single-incision.tw OR stereoscope.tw OR 3D.tw OR three dimensional.tw OR three-dimensional.tw OR NOTES.tw
Total mesorectal excision TME.tw OR total mesorectal excision.tw. OR mesorectal.tw OR rectal cancer.tw OR anterior resection.tw OR anterior-resection.tw OR TaTME.tw OR Trans-anal TME.tw OR colorectal.tw OR colo-rectal.tw. OR colo-rectal cancer.tw OR colorectal cancer.tw OR CRC.tw OR bowel resect$.tw OR rect$.tw OR proctectomy.tw
Assessment of performance performance.tw OR assessment.tw OR objective.tw OR technical.tw OR procedural.tw OR competenc$.tw OR credent$.tw OR skill.mp OR proficiency.tw OR judgement.tw OR ability.tw OR training.tw OR education.tw OR analysis.tw OR developement.tw OR task.tw OR tool.tw OR error.tw OR proforma.tw OR score.tw OR checklist.tw OR global.tw OR evaluation.tw OR validity.tw OR observational.tw
(((("laparoscopy"[MeSH Terms] OR "laparoscopy"[All Fields] OR "laparoscopic"[All Fields]) OR ("laparoscopes"[MeSH Terms] OR "laparoscopes"[All Fields] OR "laparoscope"[All Fields]) OR ("laparoscopy"[MeSH Terms] OR "laparoscopy"[All Fields]) OR keyhole[All Fields] OR (minimally[All Fields] AND invasive[All Fields]) OR minimally-invasive[All Fields] OR ("robotics"[MeSH Terms] OR "robotics"[All Fields] OR "robotic"[All Fields]) OR robot[All Fields] OR (robot[All Fields] AND assisted[All Fields]) OR stereoscopic[All Fields] OR ("squamous intraepithelial lesions of the cervix"[MeSH Terms] OR ("squamous"[All Fields] AND "intraepithelial"[All Fields] AND "lesions"[All Fields] AND "cervix"[All Fields]) OR "squamous intraepithelial lesions of the cervix"[All Fields] OR "sils"[All Fields]) OR (("single person"[MeSH Terms] OR ("single"[All Fields] AND "person"[All Fields]) OR "single person"[All Fields] OR "single"[All Fields]) AND incision[All Fields]) OR single-incision[All Fields] OR stereoscope[All Fields] OR 3D[All Fields] OR (three[All Fields] AND dimensional[All Fields]) OR three-dimensional[All Fields] OR ("Notes Czechoslov Med Assoc G B"[Journal] OR "notes"[All Fields])) AND (TME[All Fields] OR (total[All Fields] AND mesorectal[All Fields] AND excision[All Fields]) OR mesorectal[All Fields] OR ("rectal neoplasms"[MeSH Terms] OR ("rectal"[All Fields] AND "neoplasms"[All Fields]) OR "rectal neoplasms"[All Fields] OR ("rectal"[All Fields] AND "cancer"[All Fields]) OR "rectal cancer"[All Fields]) OR (anterior[All Fields] AND resection[All Fields]) OR anterior-resection[All Fields] OR tatem[All Fields] OR (trans-anal[All Fields] AND tme[All Fields]) OR colorectal[All Fields] OR colo-rectal[All Fields] OR ("colorectal neoplasms"[MeSH Terms] OR ("colorectal"[All Fields] AND "neoplasms"[All Fields]) OR "colorectal neoplasms"[All Fields] OR ("colo"[All Fields] AND "rectal"[All Fields] AND "cancer"[All Fields]) OR "colo rectal cancer"[All Fields]) OR ("colorectal neoplasms"[MeSH Terms] OR ("colorectal"[All Fields] AND "neoplasms"[All Fields]) OR "colorectal neoplasms"[All Fields] OR ("colorectal"[All Fields] AND "cancer"[All Fields]) OR "colorectal cancer"[All Fields]) OR CRC[All Fields] OR ("administration, rectal"[MeSH Terms] OR ("administration"[All Fields] AND "rectal"[All Fields]) OR "rectal administration"[All Fields] OR "rectal"[All Fields]) OR ("rectum"[MeSH Terms] OR "rectum"[All Fields]) OR proctectomy[All Fields])) AND (performance[All Fields] OR ("Assessment"[Journal] OR "assessment"[All Fields]) OR ("goals"[MeSH Terms] OR "goals"[All Fields] OR "objective"[All Fields]) OR technical[All Fields] OR procedural[All Fields] OR ("mental competency"[MeSH Terms] OR ("mental"[All Fields] AND "competency"[All Fields]) OR "mental competency"[All Fields] OR "competence"[All Fields]) OR ("mental competency"[MeSH Terms] OR ("mental"[All Fields] AND "competency"[All Fields]) OR "mental competency"[All Fields] OR "competency"[All Fields]) OR ("credentialing"[MeSH Terms] OR "credentialing"[All Fields] OR "credential"[All Fields]) OR ("credentialing"[MeSH Terms] OR "credentialing"[All Fields]) OR skill[All Fields] OR proficiency[All Fields] OR ("judgment"[MeSH Terms] OR "judgment"[All Fields] OR "judgement"[All Fields]) OR ("aptitude"[MeSH Terms] OR "aptitude"[All Fields] OR "ability"[All Fields]) OR ("education"[Subheading] OR "education"[All Fields] OR "training"[All Fields] OR "education"[MeSH Terms] OR "training"[All Fields]) OR ("education"[Subheading] OR "education"[All Fields] OR "educational status"[MeSH Terms] OR ("educational"[All Fields] AND "status"[All Fields]) OR "educational status"[All Fields] OR "education"[All Fields] OR "education"[MeSH Terms]) OR ("analysis"[Subheading] OR "analysis"[All Fields]) OR ("growth and development"[Subheading] OR ("growth"[All Fields] AND "development"[All Fields]) OR "growth and development"[All Fields] OR "development"[All Fields]) OR task[All Fields] OR tool[All Fields] OR error[All Fields] OR proforma[All Fields] OR score[All Fields] OR ("checklist"[MeSH Terms] OR "checklist"[All Fields]) OR ("Glob Impacts"[Journal] OR "global"[All Fields]) OR ("evaluation studies"[Publication Type] OR "evaluation studies as topic"[MeSH Terms] OR "evaluation"[All Fields]) OR validity[All Fields] OR observational[All Fields]) AND (("1982/01/01"[PDAT] : "2016/11/01"[PDAT]) AND English[lang])

Table 2.1: PubMed literature search string. The search was performed on November 1st, 2016.

Inclusion Criteria

Studies were eligible for inclusion if they were full-text English language articles and reported on performance assessment of minimally invasive TME resections. Studies limited to histopathological, radiological, post-operative data or non-operative factor assessment alone were excluded as well as abstracts or conference proceedings, papers containing exclusively open operations, animal studies, paediatric studies, simulated settings or case reports. Given the importance of performing a safe intracorporal anastomosis during TME surgery and the objectives of this thesis, abdominal-perineal resection (APER) cases were excluded. Randomised, minimally invasive TME trials were identified through a second search using a string previously reported (110) and cross referenced with the Cochrane database report on laparoscopic versus open TME RCTs (40). Single surgeon and/or single centre trials were excluded as they were considered to have little need to standardise or directly compare their practice with others. Trial protocols were sought from identified study websites, international research registries and personal communication with chief investigators. Two researchers independently reviewed all trial materials to identify the use, type and nature of any objective intraoperative performance assessment measure used for cases within the trials.

Data extraction

Data was independently extracted by two reviewers using a dedicated bespoke extraction form: the stated aim of each study, study design, number of patients undergoing minimally invasive TME resections, the level of both assessor and operating surgeon, what precisely was being assessed, when assessment took place and whether the assessor was blinded. Reports on validity and reliability were captured. The methodological quality of included studies was assessed independently by two authors using the Downs–Black checklist (154). This was chosen over the Cochrane risk of bias tool based upon the observed design of the majority of identified studies (155).

2.3 Results

Objective intraoperative TME assessment tools

Database searches returned 8642 citations. After abstract screening, 81 full-text publications were reviewed. Twelve papers met inclusion criteria (figure 2.1; table 2.3) (72-74, 116, 131, 133, 140, 156-160). Ten papers were prospective observational-based studies with one randomised controlled trial and one expert educational consensus. Two research groups accounted for eight of the publications with the possibility of data duplication as study dates overlapped (156, 157) or were not stated (73, 74). Included papers had a combined total of 1762 assessed cases, but just 126 (7.1%) were reported to be laparoscopic TME. Three series stated TME cases were included but did not specify exact numbers (72, 74, 116) with two papers consisting solely of TME cases (140, 158) and one also containing non-colorectal procedures (133). One tool was developed for the assessment of trans-anal TME (160), but no reports on tool use within robotic or single-incision TME surgery were identified. Only one study was based outside the UK (158). Seven papers were multicentre investigations (72-74, 131, 140, 156, 157), two of which also contained cases from the USA (156, 157). The methodological quality of the prospective, non-randomised studies was moderate according to the Downs–Black checklist (154) (median score 16, range 13–18, possible score 0–27). Wide study heterogeneity with non-comparable outcome measures and low number of TME cases precluded formal meta-analysis.

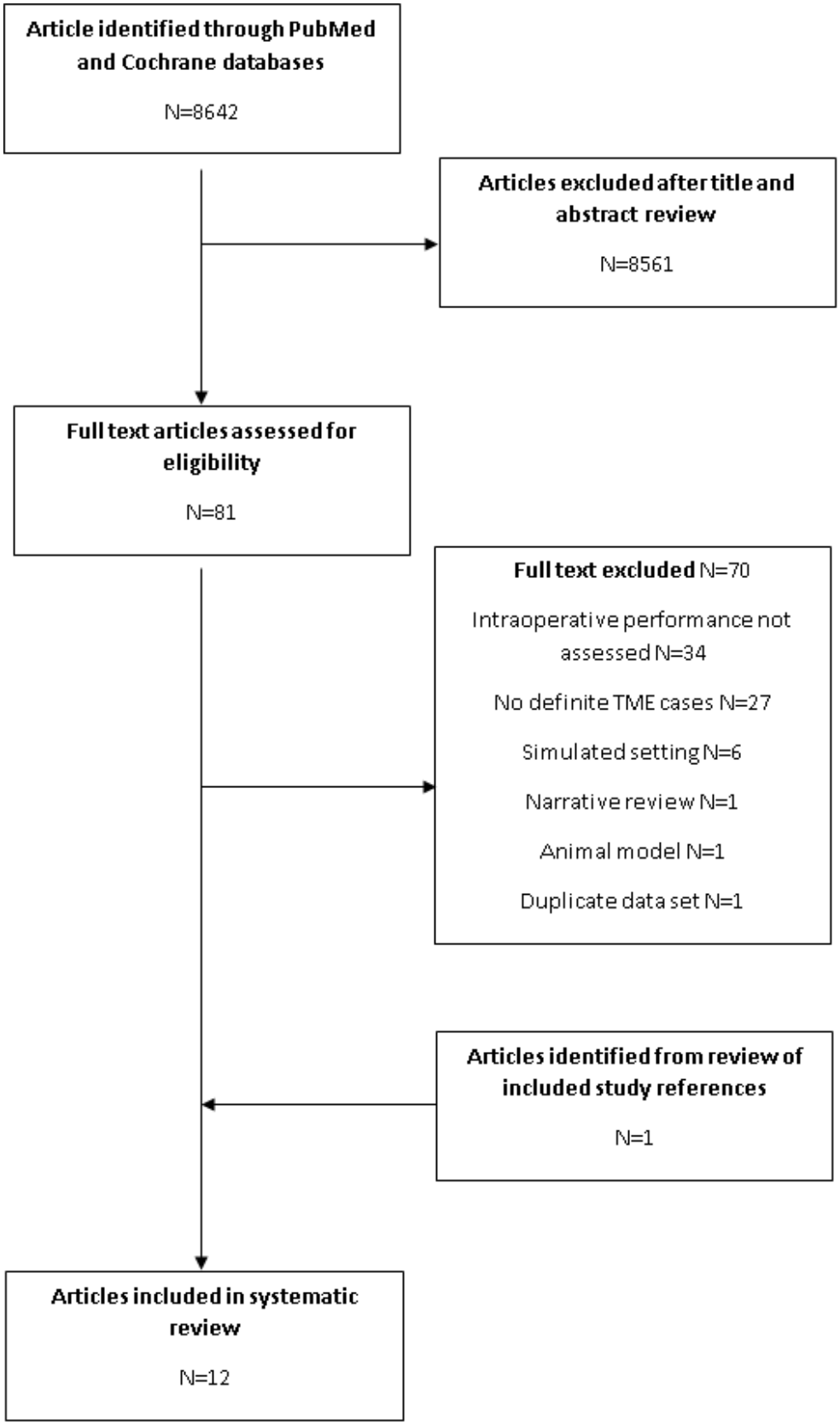


Figure 2.1 PRISMA flowchart.

Section/topic	#	Checklist item	Reported
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Chapter + paper title
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Published paper
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	2.1
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	2.2
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Not registered
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	2.2
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	2.2
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Table 2.1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	2.2
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	2.2
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	2.2 + paper supplement
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Not applicable
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Not applicable
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Not applicable

Section/topic	#	Checklist item	Reported
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Not applicable
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	Not applicable
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 2.1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 2.3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Not applicable
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	2.3. No meta-analysis
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Not applicable
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Not applicable
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	2.3+2.4
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	2.3+2.5
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	2.3+2.6
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	2.8
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Not applicable

Table 2.2: PRISMA checklist for the systematic review. Number refer to subsections of this thesis. All can additionally be reviewed in the published manuscript and supplementary materials (161).

Types of the assessment tools and methods

The included papers reported eight difference assessment tools. Two were originally created to specifically assess the intraoperative performance of laparoscopic TME (156, 160). Most were designed for colonic surgery and adapted, usually by including a TME subsection. Assessment methods were split between review of unedited operative case videos and post-operative completion of the assessment tool by the trainer or both methods (Table 2.3).

Aims of the studies and purpose of the tools

The stated aim of the included studies was to develop and validate novel self-designed assessment tools (72, 74, 156, 157, 160), validate previously described objective assessment methods (73, 133, 159), correlate intraoperative TME performance with clinical outcomes (131, 158), measure intraoperative performance within a RCT (140) and assess proficiency gain within a defined laparoscopic training programme (116). The included tools were used for a variety of purposes. Commonly, they were employed as formative assessments in both laboratory (160) and operative settings. Stated goals were the assessment of training efficacy (116), monitoring proficiency gain curves (72, 109) and competency-based assessment of procedural skill (133). Tools were also used as summative instruments for competency at the specialist level (73, 74, 131) and for generic and specific operative technical skill assessment (156, 157). Within research settings, tools were utilised to form the primary outcome measure in a multicentre RCT (140) and to investigate intraoperative TME performance (158, 159).

Formative assessment tools

Global assessment scale

The GAS tool was created by the LapCo educational committee to monitor laparoscopic colonic surgery training within their programme, but the tool was applied to rectal cases (71, 72). A generic laparoscopic colorectal resection task list was created based on review of the literature and educational materials. For each operative step, a score was assigned by the trainer dependant on the amount of support required to complete each task using a Juster scale. Sections for operative details, self-assessment, evaluation of trainer and post-operative events were incorporated into the

tool. A specific GAS form was created for trans-anal TME through expert consensus to assist the training of those taking up this technique under specialist supervision (160). This form was trialled during a cadaveric workshop, but no data was presented in the paper. Therefore, the clinical applicability, usability, validity and reliability is presently unknown.

Global modular attainment score

To assess trainees' operative performance, Jenkins and colleagues defined the individual task steps required for a laparoscopic colorectal resection (116). A weighted score representing the difficulty of that task was attached with the more steps performed by the trainee, the higher their achieved score. The amount of assistance required in their 750 cases was also documented using a modified direct observation of procedural skills scale. Additional documentation of consequential intraoperative errors was made by the supervising trainer.

Performance-based assessment (PBA)

PBA is the main procedural skills assessment method for UK surgical trainees and is a compulsory component of all programmes (133). It consists of six competency domains with generic and procedure-specific sections for rectal surgery each scored as satisfactory or unsatisfactory (or not applicable/observed) for completion of training. A global summary score of the trainee's level to perform the procedure is included alongside areas for free text from both the trainee and trainer. The tool is designed for completion immediately post-operatively by the trainer and recorded in the trainee's portfolio as a formative assessment (136).

Summative assessment tools

Laparoscopic competency assessment tool (L-CAT)

The L-CAT forms were developed for summative assessment of technical surgical performance for specialists receiving colonic surgery laparoscopic training within the LapCo programme (71). This evaluated operative competency in four main task areas (access, pedicle ligation, medial-to-lateral dissection and resection and anastomosis). The tool was applied to assess operative performance in rectal cases but does not include specific elements for the assessment of these steps (72). Content was established based on structured expert interview followed by a Delphi exercise of the

educational committee. There was unanimous agreement for a task-specific structure and descriptive scales. The L-CAT incorporates simultaneous task-specific and generic skill assessment in a hybrid 4 x 4 structure with scores of 0–4 correlating with inadequate to proficient performance (74). The L-CAT was the only tool to explore clinical and histopathological validity. Correlation with morbidity, lymph node yield and distal margin was seen although more established criteria such as circumferential margins or plane of TME surgery were not investigated (131).

Hierarchical task analysis (HTA)

Sarker broke down the steps and subtasks required for laparoscopic colorectal resections using the literature, multimedia and expert sources (156, 157). Likert scales were constructed for each procedure-specific step and overall generic skills evaluation. An anterior resection-specific HTA was created as part of this work. Independent senior surgeons assessed unedited case videos and completed the assessment. The same authors repeated this technique for surgeon self-assessment after 52 independent cases and found this was a feasible, valid and reliable strategy (157).

Research assessment tools

Observational clinical human reliability analysis (OCHRA)

Human reliability analysis was developed to evaluate the interaction between humans and complex systems in high risk industries by breaking the system down into a number of tasks where performance is prospectively analysed to identify errors and their nature and consequences. This technique has previously been adapted to complex laparoscopic procedures and termed observational clinical human reliability analysis.

Three papers utilised OCHRA as their primary method of assessment (73, 140, 159). Within the LapCo programme, this technique was used to assess competency during specialist laparoscopic colorectal training (73). The application of OCHRA to laparoscopic rectal surgery was also evaluated by two further studies investigating surgical complexity and the timing of TME surgery after completion of neoadjuvant chemoradiotherapy as well as the validity of OCHRA in TME surgery (140,

159). This was based upon blinded assessment of unedited TME case videos by reviewers trained to perform OCHRA assessment (140). Detailed OCHRA methodology is described in section 3.5.

Operative nerve injury assessment

Constanti and colleagues reviewed operative videos of patients found to have objective evidence of genitourinary disorders after undergoing laparoscopic TME surgery and directly observed evidence of iatrogenic pelvic nerve injury in half of those cases (158). All patients with post-operative dysfunction as defined using appropriate and validated questionnaires had received repeated diathermy applications for bleeding at anatomical sites that corresponded to locations where autonomic nerves would be expected to lie.

Credentialing of learners and assessors

Eight papers provided the level of the surgeon being assessed with two defining them solely by number of laparoscopic colon or rectal cases previously performed (table 2.3) (156, 157). A more detailed description of the examiners was given. Six papers used senior surgeon(s), typically defined by number of cases and/or peer-reviewed appointment. Four papers provided prior specific training to those completing the assessment measure (73, 131, 133, 140). Blinded assessment was not feasible where the surgical team were completing the assessment. All other studies were assessed by delayed independent review of anonymised case videos by assessors blinded to all events and outcomes. Although some assessment tools were based on task analysis, no paper commented on requirements to follow a set task order or whether attempts were made to standardise surgical technique. With the exception of the Constanti (158) and Foster (159) reports, operative details were often restricted to ‘‘anterior resection’’ or ‘‘TME’’ (73, 74, 131, 156, 157) with no information given on factors that could be expected to impact on the case such as patient demographics, tumour details, indication for surgery, details of neoadjuvant therapy usage or a description of their operative plan, set-up and equipment used.

Validity and reliability of the tools

Ten studies explored various validities of their assessment tool. Content validity was typically established through expert consensus and breaking down TME surgery into task step lists (72, 74, 131, 156, 157, 159, 160) and assessed using Likert (156, 157) or generic scales (72) or weighted (116) or descriptive scores (72, 74, 131, 160). Six papers assessed construct validity (72, 74, 131, 133, 156, 157) with all reporting their tool accurately differentiated between levels of performance. Clinical validity was explored by five papers. Three correlated performance with histopathological assessment (131, 140, 159) and three using post-operative morbidity (116, 131, 140). Constanti reported patient urogenital function data after TME surgery (158). All seven papers assessing tool reliability reported their instrument was reliable (73, 74, 112, 133, 156, 157, 159). Four reported inter-examiner reliability with two using generalisability theory modelling (74, 133) and two performing test–retesting (73, 159).

Paper	Design	Total cases	TME cases	Assessment tool	Assessment method	Who was assessed?	Who examined?	Credentialing of assessor	Blinded	Reliability	Validity	Timing of assessment
Sarker 2010 (156)	Observational, prospective	84	15 (18%)	Hierarchical Task Analysis	Recorded case video review	6 surgeons, 8 trainees	Two senior surgeons	>50 lap TMEs	Yes	Inter examiner	Construct	Delayed
Sarker 2011 (157)	Observational, prospective	52	4 (8%)	Hierarchical Task Analysis	Operating surgeon + assistant	4 surgeons, 6 trainees	Self-appraisal	Senior (>50 cases) or trainee (10>)	No	Inter examiner	Construct	Immediately post-operative
Miskovic 2012 (73)	Observational	32	16 (50%)	Observational clinical human reliability analysis	Recorded case video review (self-selected)	Surgeons within UK national training	Two senior surgeons	National laparoscopic trainer	Yes	Inter examiner and test-retest	Construct and concurrent	Delayed
Miskovic 2011 (72)	Observational, prospective	193	Not stated	Global Assessment Score (GAS)	Trainer + trainee completes assessment post	Surgeons within UK national training	Trainer and trainee	National trainer (>100 lap colorectal cases)	No	Inter-rater	Construct	Immediately post-operative

Paper	Design	Total cases	TME cases	Assessment tool	Assessment method	Who was assessed?	Who examined?	Credentialing of assessor	Blinded	Reliability	Validity	Timing of assessment
Miskovic 2013 (74)	Observational, prospective	54	Not stated	Laparoscopic colorectal competency	Recorded case video review	Surgeons within UK national training	Two senior surgeons	National trainer (>300 lap colorectal cases)	Yes	Generalisability theory model	Content, construct and concurrent	Delayed
Mackenzie 2015 (131)	Observational, prospective	171	6 (4%)	Laparoscopic colorectal competency	Recorded case video review	Surgeons within UK national training	Two senior surgeons	National trainer (>100 lap colorectal cases)	Yes	Not assessed	Clinical – morbidity and histopathology	Delayed
Jenkins 2015 (116)	Observational, prospective	750	Not stated	Global modular attainment score and direct	Trainer post op + Recorded case video	Eight laparoscopic fellows	Two senior surgeons	Not stated	No	Not assessed	Clinical - Morbidity	Immediately post-operative
Costanzi 2015 (158)	Observational, prospective	35	35 (100%)	Assessing for intraoperative nerve injury	Recorded case video review	Single surgeon	Full surgical team	Not stated	No	Not assessed	Clinical – Urogenital function	One and nine months
Marriott 2010	Observational, prospective	348	14 (4%)	Procedure Based Assessment	Trainer completes assessment post op	8 colorectal specialist registrars	Clinical Supervisor(surgeon)	Yes + Trained to perform assessments	No	Generalisability theory model	Face, construct and consequential	Immediately post-operative
Foster 2016	Observational, prospective	20	13 (65%)	Observational clinical human	Recorded case video review	Not stated	Research Fellow	Not stated	Yes	Test-Retest performed	Clinical - Mesorectal specimen	Delayed
Foster 2016	Randomised controlled trial	26	26 (100%)	Observational clinical human	Recorded case video review	Not stated	Research Fellow	OCHRA training and previous	Yes	Not assessed	Not assessed	Not assessed
Penna 2016	Expert consensus	Not stated	Not stated	Trans-anal TME specific Global	Trainer completes assessment	Not stated	Not stated	Not stated	No	Not assessed	Not assessed	Immediately post-operative

Table 2.3: The included studies including design features and the number, nature and aim of the performed assessments.

2.4 Application of objective assessment within multicentre laparoscopic TME randomised trials

Objective assessment of operative performance was explored within the context of the reported multicentre laparoscopic TME RCTs. Six trials met inclusion criteria (table 2.4) (83, 85, 86, 88, 89, 162), but only three (MRC-CLASSICC (83), ACOSOG Z6051 (88) and ALaCaRT (89)) assessed the intraoperative performance of laparoscopic cases. The MRC CLASSIC trial originally reported in 2005 that a laparoscopic approach was associated with similar survival, local recurrence and quality of life to those who received open TME surgery across 27 UK centres (82). A centralised random audit of case videos was performed, but no additional details were given in the publication or protocol. The American ACOSOG trial, incorporating 35 centres, failed to establish non-inferiority of the laparoscopic approach on pathological assessment of the TME specimen. A centralised audit of the first 100 laparoscopic cases was performed, but no information was provided on the nature of these assessments or their outcomes (88). The 24 site Australasian ALaCaRT RCT also failed to show non-inferiority for laparoscopic TME resection. A similar central audit was undertaken with randomly selected laparoscopic cases although no further details were provided (89). No trial reported using any tool or attempting to provide an objective assessment of intraoperative events and no study attempted to capture intraoperative performance of the open arm cases. Although challenging, capturing open rectal cancer surgery case video appears possible. The STARRCAT multi-centre RCT recorded open abdominal and perineal surgical phases via a sterile laparoscope held by a research fellow independent of the theatre team (140).

Multicentre RCT	Protocol available	Centres	lap TME cases	Was intraoperative performance assessed?	Method	How was this assessed?	Objective Assessment?	Was an assessment tool used?	Who examined?	Credentialing of assessors	Blinded?
CLASICC 2007	N	27	484	Yes - Recording of lap TME cases	Central audit of randomly lap TME case videos	NS	NS	NS	Central trial committee	NS	NS
COREAN 2014	Y	3	170	No							
EnROL 2014	Y	12	103	No							
COLOR II 2015	Y	30	699	No							
ALaCaRT 2015	Y	24	238	Yes – Recording of lap TME cases	Random audit of surgical technique	NS	No	NS	Participating hospitals	NS	NS
ACOSOG Z6051 2015	Y	35	242	Yes – Recording of lap TME cases	Central audit of first 100 videos with random case selection	NS	No	NS	Study team	NS	NS

Table 2.4: Six multicentre RCTs contained an element of objective assessment applied to included cases. Three trials recorded laparoscopic case videos, but all assessment methods appear to be subjective and unstructured. NS=not studied/stated.

2.5 Critical appraisal of results and the identified tools

This chapter aimed to report the available tools for the objective intraoperative assessment of TME performance and their prior use within multicentre rectal cancer trials. Deliberate focus was placed on intraoperative assessment in an attempt to meet the previously identified deficiencies in the application of quality assurance to this field.

In view of the concerns reported in the literature regarding TME specimen quality, it is somewhat surprising that there are few tools described to objectively assess laparoscopic TME surgery with little originally created for this purpose. Even when tools were used just 7% of cases from the included papers were TME, but all studies still felt able to conclude that objective assessment of laparoscopic TME is feasible. Therefore, surgical performance for laparoscopic rectal cancer appears underrepresented and insufficiently assessed. TME surgery appears well suited to tool based objective assessment given the importance of the singular reported technique for this complex procedure in which the anatomy and procedural steps are well defined (36, 38).

Most of the identified tools broke laparoscopic colorectal resections down into task steps and applied scoring scales to each stage of the procedure. This approach reflects the reality that complex laparoscopic procedures consist of many phases and has the advantage of identifying participants strengths as well as areas for development. Use of the global assessment scale and global modular attainment scores in their present forms do not appear appropriate for independent surgeons that are expected to be able to individual complete the cases without external assistance. These were also designed for training which risks a ceiling effect and failing to show construct validity by not sufficiently differentiating different levels of performance. Independent performance cannot be automatically assumed to imply competency and the TME case concerns identified in the RCTs were delivered by independent, credentialed, specialist surgeons.

Similarly, the use of PBA with its competent or not categories is not suited to specialist assessment and unlikely to provide insight into TME performance (118, 133, 134). All specialist surgeons would have completed their training meaning ceiling effect concerns would remain. Many features of the operative nerve injury assessment technique used by the Constanzi group appear subjective and non-reproducible (158). Although shown to be of importance to patients and post-operative quality of life (44, 45, 47), this method is too specific on one outcome and would not be expected to impact

on oncological outcomes and the suboptimal TME specimen quality shown by the major laparoscopic versus open TME trials. Therefore, the above methods are discarded for further use.

The remaining identified objective assessment methods all have promise for continuing application within this thesis. TME specific hierarchical task analyses mirror the reality of the multiple phases performed during a TME operation and is an attractive approach to base analysis. Although self-assessment and reflection is to be encouraged, its use is not appropriate for the objectives of this thesis. Likert scales enhance utility and ease of application, but numerical data alone is unlikely to reveal novel insight into the concerns raised by the laparoscopic TME trials. Statistical outliers may not be apparent as the majority of MAS TME cases obtained good outcomes. Therefore, these methods are also discarded.

The L-CAT meets many of the requirements for both an ideal assessment tool with application to laparoscopic TME cases. It has been shown to be valid, acceptable, reliable and usable in specialist laparoscopic assessment with an associated educational impact and negligible cost implication. The primary drawback is that this instrument was designed for colonic cases and although the paper states it has been applied to TME cases no further data was provided (74). Direct translation outside colonic surgery particularly for high stakes assessment is not appropriate without dedicated study.

The OCHRA technique appears to meet all requirements for ongoing use in this research. Through detailed review of the entire intraoperative period, error events and importantly the context behind them gives the highest level of detail from all the identified tools. Information on the factors, timing, instrument, location and mechanism behind each event could shed new insights into TME surgery both where it is delivered with good outcomes and also highlight problem areas that could be recognised and targeted for improvement. The primary drawback to OCHRA use is the long analysis time required but this remains feasible within the research setting.

Few included studies defined a threshold between safe and unacceptable performance. Objective assessment tools have been successfully used to enhance learning and for signing off in the LapCo

programme, but these were designed to assess training on laparoscopic colonic surgery and not TME dissection as this fell outside their remit. Therefore, it would be logical to expect that similar tools are required to objectively assess performance in minimal access TME to ensure quality assurance of the surgical intervention before independent practice. As none of the identified tools meet all requirements for the objective assessment of high stakes specialist performance in routine clinical practice, bespoke tool design appears indicated.

2.6 Addressing the identified need

Features of an ideal TME assessment tool

None of the identified tools met all criteria required for an ideal instrument. Specialist performed task assessment would therefore be further served by the development of a novel, dedicated TME-specific objective tool. Inclusion of all TME dissection and pelvic tasks and the ability to differentiate between high and low performance levels form key validity requirements. Inter- and intra rater reliability should also be considered essential prerequisites to adoption. Tool usability facilitating timely administration and prompt feedback should be considered. Identified studies used either immediate or delayed tool completion. An ideal tool should therefore be equally applicable in both settings. Whilst the tool should primarily aim to assess surgical performance, clinical and histopathological validity are important requirements given their clear link with patient and oncological outcomes (20, 77, 101).

The laparoscopic TME performance tool (unpublished work)

Prior recognition of this need led to research aiming to design such a tool taking place within Imperial College. A novel laparoscopic TME performance tool (L-TMEpt) was previously developed by my predecessors, Jake Foster and Danilo Miskovic with the full report presented in their theses (figure 2.2) (129). Their following summary is included to provide an understanding of the tool.

"Through structured review of laparoscopic TME case videos together with review of the international TME standardisation report which was based on a 60 expert surgeon Delphi exercise,

the main constituent task steps were deconstructed (38). Abdominal phases of the procedure were not included as they can be assessed the validated L-CAT instrument (74).

Objective descriptors of technical performance were developed through analysis of interviews with the same 60 expert surgeons and further structured evaluation of unedited laparoscopic TME case video. These were mapped into four domains of performance which describe different technical aspects within each task area. Further refinement took place at an interactive expert workshop.

Performance assessment metrics were developed with experts in laparoscopic colorectal surgery who were invited to participate in a short interview conducted either in person or by telephone. The experts were identified through their experience in laparoscopic TME, on peer recommendation or their involvement in clinical trials investigating laparoscopic colorectal surgery. All experts participated on a voluntary basis without payment.

A semi-structured interview framework was used allowing freedom to express their thoughts and explore ideas, whilst also enabling the interviewer to cover all necessary information (163). Open questions were used to determine what indicators of performance the expert would look for to assess laparoscopic TME surgery technical performance. Additionally, for each task area, two short video clips were prepared and shown with the expert asked to reflect upon the performance.

Interviews were transcribed verbatim and analysed using qualitative thematic analysis coding methods. Thematic saturation was considered to have been achieved following the eighth interview. Descriptors of both proficient and poor performance were collated from the transcripts and triangulated into the specific procedural tasks to generate the assessment metrics for the tool.

In addition, common technical error mechanisms seen in laparoscopic rectal resection were identified from an experiment in which the OCHRA technique was applied to video recordings of laparoscopic rectal cancer resection procedures (159). Described technical errors informed the errors domain.

A draft assessment form was generated with four task areas reflecting steps of the procedure described in the expert consensus. A four-point scale was used to describe the quality of technical performance for each domain within each task area: these objective descriptors were developed from the interviews and error analysis and refined through discussions amongst the steering group. In addition, the complexity for each case was scored on the assessment form using a three-point scale (1=Wide pelvis; no scarring/oedema; no gross obesity, 2=Moderate width pelvis; minimal scarring/oedema in tissue planes; moderate bulk to tumour/mesorectum, 3=Narrow pelvis; significant scarring/oedema/reaction to neo-adjuvant therapy; bulky tumour; obesity). To avoid cluttering, an instruction manual was developed to accompany the score sheet (129).

The draft tool was piloted on a series of twelve prospective laparoscopic TME cases. Anonymised, unedited videos were independently assessed by two of the eight national trainers using the tool. Scores ranged from 34 to 57 (possible range 16-64). There was a moderate significant inverse correlation between overall score and procedure complexity grading (r^s= -0.504, p=0.008). Internal consistency across scoring items was high (Cronbach's alpha = 0.879) with strong inter rater reliability observed (ICC=0.805, p=0.017). It was concluded that the L-TMEpt satisfies many of the requirements of a useful assessment tool with potential application for formative and summative assessment and intraoperative quality assurance".

Development of the L-TMEpt

The tool development work has been reported (129). The L-TMEpt represents a bespoke tool for assessment of technical performance of laparoscopic TME surgery and as such appears to successfully meet my requirements. However, the L-TMEpt now requires completion and development through a larger scale evaluation including validity and reliability assessments. This would be possible through adoption for ongoing use within this thesis and is added as objectives.

Figure 2.2 (next page): The laparoscopic TME performance tool. Figure originally presented in Mr Jake Foster's Imperial College London thesis held in the Spiral Repository (129). Reproduced under a Creative Commons Attribution Non-Commercial No Derivatives licence and with the author's (copyright holder) permission.

L-TMEpt	1. Posterior Mesorectal Dissection	2. Anterior Mesorectal Dissection	3. Lateral Mesorectal Dissection	4. Resection & Anastomosis
Retraction and Exposure	**Retraction and exposure throughout task:**	**Retraction and exposure throughout task:**	**Retraction and exposure throughout task:**	**Preparation of bowel for resection, view of pelvis:**
	☐ **4.** Clearly demonstrates all landmarks. Optimal traction and tissue tension throughout.	☐ **4.** Clearly demonstrates all landmarks. Optimal traction/counter-traction & tissue tension.	☐ **4.** Clearly demonstrates planes. Optimal traction/counter-traction & tissue tension.	☐ **4.** Optimal distal mobilisation of rectal stump. Clear views throughout anastomosis.
	☐ **3.** Demonstrates most landmarks; appropriate traction and tension on tissue.	☐ **3.** Demonstrates most landmarks. Appropriate traction/counter-traction & tension.	☐ **3.** Good demonstration of planes. Appropriate traction/counter-traction & tension.	☐ **3.** Adequate mobilisation of rectal stump. Adequate views of pelvis during anastomosis.
	☐ **2.** Ineffective demonstration of landmarks; traction often in wrong direction. Little tension.	☐ **2.** Ineffective demonstration of landmarks. Poor traction/counter-traction. Little tension.	☐ **2.** Ineffective demonstration of planes. Sub-optimal traction/counter-traction. Little tension.	☐ **2.** Distal stump not fully mobilised. Sub-optimal views deep in pelvis during anastomosis.
	☐ **1.** Fails to demonstrate landmarks. Poor views & traction. Closed tissue planes. No tension.	☐ **1.** Fails to demonstrate landmarks. Poor views & traction/counter-traction. No tension.	☐ **1.** Fails to demonstrate correct plane. Poor views & traction/counter-traction. No tension.	☐ **1.** Distal stump tethered. Poor views deep in pelvis during anastomosis.
	☐ Unable to comment	☐ Unable to comment	☐ Unable to comment	☐ Unable to comment
Task Performance/ Execution	**Dissection in posterior TME plane:**	**Dissection in anterior TME plane:**	**Dissection in right and left lateral TME planes:**	**Placement and use of staplers (linear & circular):**
	☐ **4.** Optimal dissection in correct plane throughout. Safe efficient instrument use.	☐ **4.** Peritoneal reflection incised at optimal site. Correct plane followed. Efficient movements.	☐ **4.** Dissection follows established planes. Clear bloodless plane developed safely & efficiently.	☐ **4.** Proficient linear stapler placement; 1-2 cartridges used. Efficient use of circular stapler.
	☐ **3.** Occasional loss of plane quickly corrected. Atraumatic tissue handling & instrument use.	☐ **3.** Appropriate incision site. Quickly corrects loss of plane. Safe instrument use.	☐ **3.** Appropriate plane developed safely with minimal bleeding. Quickly corrects plane loss.	☐ **3.** Safe stapler use, but needing adjustments. 1-2 cartridges used.
	☐ **2.** Ineffective dissection. Repeated loss of plane. Inefficient, laborious instrument use.	☐ **2.** Sub-optimal incision site. Repeated loss of plane. Inefficient, laborious instrument use.	☐ **2.** Unclear relationship to established planes. Dissection into fat. Inefficient instrument use.	☐ **2.** Suboptimal stapler manipulation and placement. 3 cartridges or extra clip needed.
	☐ **1.** Uncorrected dissection in wrong tissue plane. Blunt tissue injuries. Dangerous instrument use.	☐ **1.** Incorrect site for incision. Mostly in wrong or unclear plane. Dangerous instrument use.	☐ **1.** Dissection not in a defined plane or in wrong plane. No attempt to follow established planes.	☐ **1.** Dangerous, forceful stapler manipulation & tissue handling. Multiple cartridges needed.
	☐ Unable to comment	☐ Unable to comment	☐ Unable to comment	☐ Unable to comment
Errors	**This task was performed with:**	**This task was performed with:**	**This task was performed with:**	**This task was performed with:**
	☐ **4.** No bleeding/avulsion. Hypogastric nerves safeguarded. No collateral injury/perforation.	☐ **4.** No bleeding/avulsion Neurovascular bundles safeguarded. No collateral injury/ perforation.	☐ **4.** No bleeding/avulsion. Inferior hypogastric plexuses safeguarded. No collateral injury.	☐ **4.** Ideal compression. Stapler pin enters at optimal site. Clearly confirms no twist/tension.
	☐ **3.** Minimal bleeding/avulsion. Risk unilateral nerve injury. No collateral injury/perforation.	☐ **3.** Minimal bleeding/avulsion. Risk unilateral Ny bundle injury. No collateral injury/perforation.	☐ **3.** Minimal bleeding/avulsion. Risk unilateral plexus injury. No collateral injury/perforation.	☐ **3.** Adequate compression. Pin enters at appropriate site. Checks for twist/tension.
	☐ **2.** Moderate bleeding/avulsion. Likely unilateral nerve injury. Risk of collateral injury.	☐ **2.** Moderate bleeding/avulsion. Likely unilateral Ny bundle injury. Risk of collateral injury.	☐ **2.** Moderate bleeding/avulsion. Likely unilateral plexus injury. Risk of collateral injury.	☐ **2.** Insufficient compression. Stapler pin offset. No check for twist/tension/interposed tissue.
	☐ **1.** Substantial bleeding/avulsion. Both nerves probably divided. Collateral injury/perforation.	☐ **1.** Substantial bleeding/avulsion. Bilateral Ny bundle injury. Collateral injury/perforation.	☐ **1.** Substantial bleeding/avulsion. Bilateral plexus injury. Collateral Injury/perforation.	☐ **1.** No compression. Twist/ tension not checked. Tissue involved in anastomosis. Perforation.
	☐ Unable to comment	☐ Unable to comment	☐ Unable to comment	☐ Unable to comment
End-product	**Quality of mesorectum and pelvis after task:**	**Quality of mesorectum and pelvis after task:**	**Quality of mesorectum and pelvis after task:**	**Quality of rectal stump and anastomosis:**
	☐ **4.** Smooth intact bi-lobar mesorectum. No fascial injury. No mesorectal tissue left in situ.	☐ **4.** Smooth anterior mesorectum, +/- intact shiny Denonvillier fascia. No tissue left in situ.	☐ **4.** Smooth lateral surface of mesorectum with no defects. No mesorectal tissue left in situ.	☐ **4.** Optimal transection level. Staple lines in continuity. Optimal,well-perfused anastomosis.
	☐ **3.** Occasional minor mesorectal injury to fascia only. Minimal tissue left in situ in pelvis.	☐ **3.** Occasional injury to mesorectal/Denonvillier fascia. Minimal tissue left in situ in pelvis.	☐ **3.** Occasional minor mesorectal injury to surface only. Minimal tissue left in situ.	☐ **3.** Good transection at appropriate level. Staple lines meet, mostly straight. Safe anastomosis.
	☐ **2.** Sub-optimal specimen with injury into fat. Some mesorectal tissue remaining in pelvis	☐ **2.** Sub-optimal specimen with injury into fat. Some mesorectal tissue remaining in pelvis.	☐ **2.** Sub-optimal lateral specimen with injury into fat. Some mesorectal tissue still in pelvis.	☐ **2.** Zig-zag/dog-leg transection line. Anastomosis probably safe, but possible tension.
	☐ **1.** Incomplete posterior mesorectum with deep injuries into fat/to rectum. Tissue left in pelvis.	☐ **1.** Incomplete anterior aspect to specimen. Deep fat injuries/to rectum.Tissue left in pelvis.	☐ **1.** Incomplete lateral aspect to specimen. Deep fat injuries/to rectum. Tissue left in pelvis.	☐ **1.** Staple lines do not meet. Tension/twist/poor perfusion of anastomosis.
	☐ Unable to comment	☐ Unable to comment	☐ Unable to comment	☐ Unable to comment

Complexity:

☐ **1.** *Wide pelvis; no obvious scarring/oedema in tissue planes; small tumour; mesorectum not bulky*
☐ **2.** *Moderate width pelvis; minimal scarring/oedema in tissue planes, moderate bulk to tumour/mesorectum*
☐ **3.** *Narrow male pelvis; marked scarring/oedema; bulky tumour/mesorectum*

2.7 Selection of tools for application to rectal cancer trials

To explore my hypotheses two tools have been selected for ongoing use in this thesis for application to multi-centre laparoscopic TME trials – OCHRA and the laparoscopic TME performance tool.

OCHRA, based upon the post-operative review of unedited case video, has been successfully applied to and validated for the assessment of intraoperative specialist laparoscopic TME performance including as the primary endpoint of randomised controlled trials. This approach allows a blinded, objective assessment of the entire intraoperative period. This methodology will capture all events including non-consequential errors that would otherwise be ignored (132, 164-166). This risks losing an important data source for quality improvement and patient safety and goes against established practices of comparable high risk industries (167, 168). Time and resource requirements realistically limit OCHRA to research settings, but the detail obtained justifies its adoption in this thesis. OCHRA can provide insight and detailed analysis on intraoperative errors that can be correlated with patient outcomes as has been shown in other complex laparoscopic procedures (169-173). This strategy may identify commonly observed errors which can be targeted with focussed training initiatives.

The L-TMEpt represents a bespoke tool created specifically for this purpose. The developmental methodology behind the L-CAT which was successfully utilised for high stakes specialist assessment of laparoscopic colonic surgery has been reproduced for TME procedures (129). The short time to complete at the end of the case by the assessor is appealing and is likely to assist clinical utilisation. It can be applied by an assessor in the operating theatre or following video case review which would also allow simultaneous OCHRA assessment.

This combination may be advantageous as the techniques appear complementary and should provide a blend of research depth and real-world utility in keeping with the goals of this thesis. However, a theoretical risk that OCHRA and L-TMEpt are insufficient to explore the hypotheses of this thesis exists. The time intensive nature of OCHRA may prove a barrier to future routine clinical use if shown to be of value. Whilst the bespoke design of the L-TMEpt meets requirements, it has

not yet been validated or has its reliability been established. Therefore, these key aims are included in my hypothesis and objectives and investigated in the following chapters.

The systematic review performed has a number of limitations. The search was limited to reported TME specific assessment tools with application to laparoscopic cases. Although comprehensive search terms were used, it is possible that existing tools, including generic instruments or those used in open surgery, were overlooked. These may have been applicable, given that the same dissection is performed irrespective of surgical approach and chosen instruments. Abdominoperineal cases were excluded to avoid variability from differing surgical approaches and anatomical planes including when the perineal dissection is undertaken (174). Safe intracorporal anastomosis is a fundamental part of laparoscopic TME surgery but not for APER cases. Again, as the TME dissection is identical with same requirements for high quality surgery there is a risk of missing applicable tools which may be of particular relevance given the challenges of low rectal cancer resection. As only two included studies reported case details (131, 158), it is not possible to comment on important relevant factors such as neoadjuvant therapy use, tumour stage and height from the anal verge. Case selection bias cannot be excluded in the identified reports particularly where surgeons self-selected operative videos to submit for analysis. Meta-analysis was not possible owing to the wide study heterogeneity with non-comparable methodological approaches and outcome measures identified.

2.8 Conclusions

Objective assessment of laparoscopic total mesorectal excision performance is possible but has rarely been utilised for intraoperative quality assurance purposes. A systematic review of the literature identified a number of tools that have been used for this purpose. After review and critique, the objective clinical human reliability analysis and the bespoke laparoscopic TME performance tool, created in response to the identified deficiencies in the reported literature, are selected for ongoing use in this thesis and exploration of my hypothesis through development and application to laparoscopic TME surgery including cases within RCTs.

Chapter 3 – Testing the application of intraoperative quality assurance to a laparoscopic total mesorectal excision randomised trial

The selected intraoperative assessment tools now require application to laparoscopic TME cases.

I am hugely grateful for the support from the Australasian Laparoscopic Cancer of the Rectum trial management group in particular Associate Professor Andrew Stevenson (University of Queensland, Brisbane) and Prof. Peter Hewett (University of Adelaide) for making this collaboration possible. The National Health and Medical Research Council of Australia clinical trials centre (University of Sydney) were wonderful hosts and provided the matched anonymised clinical data for analysis. My thanks go to Mr Chris Brown (biomedical statistician, University of Sydney) for his expert assistance. I am grateful for the introduction to the Australian group made through Prof Tan Arulampalam (Colchester). The fantastic opportunity to visit Australia was made possible by an Imperial Dean's internship award which I was honoured to receive.

Miss Sarah Abbott (Consultant Surgeon, Canterbury Health Board, New Zealand) acted as a second independent assessor of the entire ALaCaRT video series during her colorectal fellowship training under Prof Hewett. This was a significant extra undertaking and greatly assisted the reliability and validity assessments performed.

3.1 Introduction

The previous chapter identified that quality assurance through intraoperative performance assessment is possible but has been rarely performed within laparoscopic TME surgery. OCHRA and the L-TMEpt objective assessment techniques therefore now require application to the intraoperative period. In view of the robust methodology and high quality clinical data associated with clinical RCTs it is desirable to study cases from trials. Together with their associated high profile and high level, practice shaping evidence provided, application to RCTs may help contextualise their results and offer new insight into laparoscopic TME surgery including areas for improvement. As RCTs form the criterion standard establishing QA reliability and validity using their data would represent a strength and assist further applications of these techniques.

Aim

In view of my hypotheses the aim of chapters 3-5 is to develop and apply the identified QA tools to surgical cases from laparoscopic TME randomised trials and assess their ability to reliably assess operative technique and clinical outcome validity.

3.2 Trial selection for surgical performance assessment

As shown in table 2.4 three multicentre trials captured TME case videos potentially allowing application of intraoperative quality assurance processes and comparison with RCT clinical data. Trial summaries including their inclusion criteria are displayed below.

Trial	MRC CLASICC (82, 83)	ALaCaRT (89)	ACOSOG Z6051 (39, 88)
Recruitment period	1996-2002	2010-2014	2008-2013
Location	UK	Australia and New Zealand	USA
Centres	27	24	35
Total trial patients	794	475	486
Randomised to lap TME	253	238	243
Conversion	34%	9%	11%
Per protocol lap TME (n)	160	217	213
Surgeon credentialing	≥20 lap assisted resections	100< lap colectomies and 30< lap TMEs with verification Pre-trial male lap TME case video review	≥20 lap or hand assisted TME resections with verification One pre-trial lap TME case video review
Patient/TME inclusion criteria	Not stated	≤15cm from verge Life expectancy ≥12 weeks Age 18+	Age 18+, BMI <34 ≤12cm from verge cT3N0-2M0
Patient/TME exclusion criteria	≤5cm from the anal verge Poorly differentiated Synchronous cancers Other concurrent gastrointestinal pathology	cT4 or involved CRM Concurrent or past pelvic malignancy	cT4 ASA IV-V Previous laparotomy or pelvic malignancy
Neoadjuvant treatment	Not stated	As per local centre practice	Mandatory
Intervention details	Surgeons current practice Abdominal extraction site	Not dictated. Surgeons usual practice.	Hybrid approaches allowed (open abdomen/lap TME) All splenic flexures mobilised Standard vessel and mesentery protocol

Table 3.1: The three completed multicentre laparoscopic vs. open TME RCTs. Trial summaries including surgeon and patient eligibility data.

This data shows that the CLASICC and ACOSOG Z6051 trials appear unsuitable for investigating my hypotheses. As suggested by the MRC-CLASSIC recruitment period and very high conversion rate it is questionable if data from the study is applicable to contemporary laparoscopic TME practice. Additionally, advances in staging, multi-disciplinary working, laparoscopic technologies and most importantly surgeon experience and training could be argued to limit the applicability of the results to current practice and contemporary areas of concern. 20 cases falls far below the reported proficiency curve and whilst may have been appropriate at the time of trial design, that figure is below current requirements for completion of colorectal subspecialty training (56). Many UK

surgeons would not routinely perform abdominal-perineal resections for tumours around the 5cm height cut off used. These low cancers are of particular interest given the surgical challenges and higher rates of margin involvement reported (175, 176). Overall, analysis of this trial risks being uninformative to contemporary laparoscopic TME surgeons and the concerns raised by recent trials.

The American ACOSOG Z6051 trial holds a number of strengths but there are concerns regarding generalisable to UK practice. In keeping with US guidelines all patients underwent neoadjuvant chemoradiotherapy for cT3N0-2 disease whereas UK clinicians would typically limit pre-operative treatment to potentially involved circumferential resection margins (15, 30, 146). Hybrid procedures where the TME and pelvic surgical tasks are performed according to randomisation with the abdominal phase open to surgeon preference was allowed. This would be likely to reduce the number of complete videos available for review and introduce significant procedural heterogeneity with variations in early postoperative data. Although this would be partially offset the mandatory splenic flexure mobilisation requirement, this may also reduce validity in UK surgery as this practice is reported to vary between surgeons and individual case requirements (177).

The Australasian Laparoscopic Cancer of the Rectum trial (ALaCaRT) appears well suited to my objectives. The inclusion criteria reflect UK practice for neoadjuvant treatment use. The higher surgical experience requirements also increase the comparative strength with UK consultant surgeon experience but remain under numbers reported for completion of laparoscopic TME proficiency curves (56). 7.5% of the laparoscopic arm underwent APER surgery slightly reducing the number of anastomoses expected in the video series and 14% of cases were performed robotically which would also need to be excluded. However, their large number of laparoscopic resections should allow extensive application of objective intraoperative performance assessment.

3.3 ALaCaRT trial synopsis

Detail	Description
Design	Multicentre, randomised, phase III, non-inferiority trial
Aim	To determine whether laparoscopic resection is noninferior to open rectal cancer resection for adequacy of cancer clearance
Primary endpoint	Composite endpoint of adequate surgical resection consisting of 1) Complete TME – mesorectal plane of surgery 2) Clear circumferential margin (≥1mm) 3) Clear distal resection margin (≥1mm)
Randomisation	Central randomisation upon recruitment (1:1) Stratified by tumour height, surgeon, planned procedure, body mass index, neoadjuvant treatment and presence of distant metastases
Intervention	Laparoscopic total mesorectal excision
Control	Open laparotomy and TME
Location	26 credentialled surgeons from 24 Australasian centres
Outcome assessment	Blinded histopathologists following standardised assessment protocol
Sample size	470 patients required to determine laparoscopy as noninferior with a margin of Δ -8% (80% power, α 0.05) assuming 90% surgical success rate and conversion rate less than 20%
Study Registry	Anzctr.org identifier ACTRN120609000663257
Ethical approval	Sydney local health district human ethics research committee
Chief Investigator	Associate Professor Andrew Stevenson, University of Queensland
Clinical Trials Centre	NHMRC Clinical trials Centre, University of Sydney
Funder	Colorectal Surgical Society of Australia and New Zealand
Sponsor	Australasian Gastro-Intestinal Trials Group

Table 3.2: Synopsis of the ALaCaRT study.

The full ALaCaRT protocol is available as a supplement to the trial publication (89). Regarding the capture of laparoscopic TME case video, the monitoring of surgical performance section states "*video audit of laparoscopic procedures will take place throughout the trial with random assessment of 20% of submitted videos after accrual of 50, 150, 250 and 350. Therefore, all laparoscopic procedures must be filmed and confirmed to have been filmed on the appropriate case report form. Failure to film procedures will be noted by the study chair and reported to the trial management committee. The study chair and management group are empowered to suspend protocol participation if necessary, until appropriate facilities to film laparoscopic procedures are in place. (protocol 7.4.2)".* Therefore, the ALaCaRT study was considered ideal for the application of intraoperative quality assurance within laparoscopic TME surgery due to the methodological

strengths, size and quality assurance techniques used. Performance assessment has the potential to provide insights into the mechanisms behind the trial results.

3.4 Contact and sub study proposals and approvals

Introductions were kindly made by Prof Tan Arulampalam (Consultant colorectal surgeon, Colchester) to Prof Peter Hewett (University of Adelaide) and the ALaCaRT chief investigator Associate Professor Andrew Stevenson (University of Queensland). After initial fruitful discussions and provisional agreement to collaborate a formal proposal entitled “Reliability and Validity Evaluation of the Laparoscopic TME Objective Assessment Tool: ALaCaRT sub study proposal” was submitted to the ALaCaRT trial management group for review. After acceptance and completion of contractual arrangements, I was welcomed to Australia in October 2018 visiting both the chief investigator and national clinical trials unit at the University of Sydney.

3.5 Methods

Objectives

1) Complete training in OCHRA methodology and its application to laparoscopic TME case video
2) L-TMEpt reliability evaluation
    a. Inter-rater L-TMEpt reliability
    b. Test-retest (intra-rater) reliability
3) L-TMEpt concurrent validity
    a. Application of the Objective Clinical Human Reliability Analysis (OCHRA) method
    b. Comparison of OCHRA and L-TMEpt results
4) L-TMEpt clinical validity
    a. Histopathological specimen analysis – plane of surgery, clear margins

b. Early morbidity and short term patient outcome data

c. Oncological outcomes - survival and recurrence data

5) Establish a clinically valid L-TMEpt threshold score for acceptable performance

OCHRA training

Before undertaking independent assessments a period of OCHRA methodology training was undertaken. After reviewing the available literature, I received OCHRA training from a surgical researcher experienced in performed OCHRA assessments for laparoscopic rectal cancer surgery (Mr Jake Foster, Imperial College). Initial teaching was didactic before observing live OCHRA assessments of previously analysed unedited laparoscopic TME videos from the completed STARRCAT RCT (ISRCTN88843062) (140). Inter-rater reliability was explored using the total number of case errors. Joint analysis of three further cases with comparison and discussion of results including personal thresholds for error scoring and classification was performed. Next, I independently reviewed the available series of 20 consecutive, unedited laparoscopic TME videos previously captured by Mr Foster & Prof Francis at Yeovil District Hospital NHS Foundation Trust between 2012 and 2013 with all previous error coding events and details removed (159). As a summative assessment, Prof Francis and Mr Foster observed me performing an independent TME case analysis. Test-retest intra-rater reliability of three randomly chosen cases from the TME series was performed after a three-month delay. Video files were re-coded by an independent research staff member to maintain blinding and prevent comparison with prior results.

Reliability evaluation

The reliability of the L-TMEpt was investigated through scoring from two independent assessors with laparoscopic TME experience, blinded to all clinical and specimen data. Reliability was initially assessed through calculation of the interclass correlation coefficient and Cronbach alphas statistic. Total scores were analysed as a continuous variable with subsequent exploration of the score for each individual item, skill domain and task objective (see tool scoresheet figure 2.4). Inter-rater reliability of the assessor case complexity score was reported using Cohen's kappa (κ, grade 1 easiest

- 3 hardest). Test-retest reliability was explored through full re-analysis of the entire video series at least 12 months after original review using the L-TMEpt with ICC and Cronbach's alpha calculated with 95% confidence intervals.

Concurrent validity evaluation

As identified in chapter two, the observational clinical human reliability analysis (OCHRA) is a technique developed to assess the interface between humans and complex systems; whereby the system is described in its constituent tasks with each task analysed to identify potential errors and their consequences in a prospective manner. This was initially utilised by high risk industries and has been successfully modified for the detection and categorisation of technical errors and near-misses enacted within laparoscopic surgical procedures (73, 159). The insight provided can help highlight technical challenges, particularly within complex procedures such as TME.

Development of OCHRA methodology for laparoscopic TME

Laparoscopic TME procedures were broken down into constituent task steps using previously reported laparoscopic TME specific hierarchical task analyses (table 3.3) (38, 156). OCHRA categories were based upon previous descriptions used in TME and other laparoscopic procedures (73, 159, 170-172, 178-180) and refined using TME specific journal articles, descriptions in surgical textbooks, and available laparoscopic rectal cancer operation case video. This was used to create a schedule for a series of semi-structured interviews with leading experts that helped define the international structured laparoscopic TME consensus (38). These transcripts were additionally used to create TME specific error, consequence and instrument categorical lists (tables 3.6).

Unedited videos were reviewed maintaining assessor blinding to all clinical details including surgeon, centre and date of recording. Files were managed using Adode Premiere Pro (v6.0, Abode, San Jose, CA, USA). In view of previous reports showing errors were seen to take place across all TME tasks, a patient centred approach was adopted with analysis of the entire operation (140, 159). Error events were defined *"as something done which was not intended by the actor nor desired by a set of rules or an external observer or led the task outside acceptable limits"* (168). Identified events were

further categorised by timepoint, operative task phase, instrument used, external error mode, instrument/dissection or tissue/retraction errors (based upon the perceived principal mechanism for the event) and any resulting consequence using previously reported pre-defined coding lists (tables 3.3 and 3.6). Any error taking part in the pelvis was assigned an additional code based on the spatial location of the event using a previously reported distribution (figure 3.1) (140). Deviations in task sequence were not considered errors.

The overall sum of total case errors and the number of pelvic error events were used to assess the construct validity of the L-TMEpt total score using the Wilcoxon signed-rank and Spearman rho tests. Construct validity of OCHRA was assessed through TME specimen quality defined as the histopathologically reported mesorectal plane of surgery (14) (table 4.1).

Table 3.3 (next page): The hierarchical task analysis applied to this study. Adapted from the international laparoscopic TME standardisation project (38). *Not captured by video recordings. †Optional tasks that may not be performed in all procedures (177).

Task	Subtasks	Start point	Aim	Expert consensus
1 Preparation of patient*	Safe positioning on table Preparation of abdomen	Removal of patient blankets	Patient safely positioned. Abdomen prepared and draped	93%
2 Exposure of operating field	Port insertion & pneumo-peritoneum Preparation of operating field	Knife-to-skin Grasping of small bowel / mesentery / omentum	Ports in ergonomic location. Pneumoperitoneum established Small bowel mobilised out of pelvis and operating field. Adhesions divided. Pedicle exposed	93%
3 Division of vascular pedicles	IMA pedicle IMV division	Incision of peritoneum over sacral promontory Division of peritoneum over IMV	IMA identified, skeletonised and divided IMV secured + transected at appropriate level	96%
4 Mobilisation of descending colon	Medial to lateral mobilisation of descending colon mesentery Lateral mobilisation	Mesentery released from Toldt's fascia Retract colon medially	Descending colon mobilised Left ureter & gonadal vessels safely preserved	100%
5 Mobilisation of splenic flexure†	Medial mobilisation Inferior mobilisation	Incision of peritoneum between transverse colon and greater omentum Mobilisation of mesentery overlying pancreas	Transverse colon and splenic flexure fully mobilised	100%
6 Posterior TME	Posterior mesorectal dissection Right and left posterolateral mesorectal dissection	Recto-sigmoid junction retracted anteriorly to open up posterior holy plane.	TME plane developed posteriorly without fascial breach. Hypogastric nerves preserved Mesorectal plane extended to left and right posterior-lateral aspects. Avoid ureteric and pelvic sidewall nerve injury	95%
7 Anterior TME	Anterior mesorectal dissection	Incision of peritoneum of anterior reflection	Rectum released from anterior relations without fascial injury. Denonvillier's fascia divided distally	100%
8 Distal TME	Distal posterior TME Distal anterior & lateral TME	Mesorectal dissection continued to termination Rectum retracted posteriorly	Exposing rectal tube muscle Lateral attachments of rectum divided, and muscle tube exposed circumferentially	95%
9 Division and anastomosis	Cross-stapling of rectum Evacuation of specimen* Anastomosis	Stapler visible Bulge of rectal stump from tip of circular stapler	Divided with 1 or 2 firings with staple line at right angles to the rectum Safely anastomosis. Not under tension and no twist. Air leak test	86%
10 Completion†	Drain placement† Stoma formation†	Appearance of drain on screen Grasping of terminal ileum	Drain(s) placed Loop ileostomy created	91%

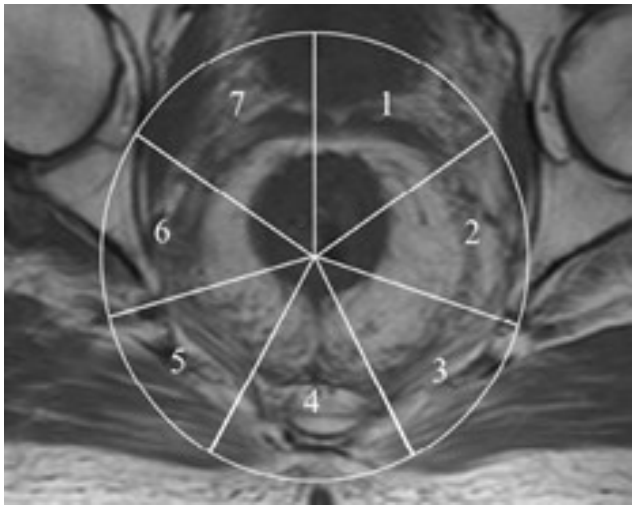


Figure 3.1. Pelvic zones used to group error locations. Reproduced from Foster et al (159) with written permission (appendix 1).

L-TMEpt assessment tool

In addition to the application of OCHRA to the ALaCART videos, L-TMEpt clinical validity was tested through analysis against operative, specimen, short term morbidity (using the Clavien-Dindo classification (181-183)) and oncological outcome data. Cases were analysed using the ALaCaRT composite histopathological endpoint and mesorectal fascial plane of surgery before attempts to define thresholds of clinical significance using L-TMEpt score as a continuous variable. OCHRA case error counts were compared with L-TMEpt scores using spearman's rho correlation co-efficient to assess the nature and strength of any association.

Establish a L-TMEpt score (threshold) cut off

An attempt to quantify an L-TMEpt total "threshold" score associated with both acceptable and suboptimal clinical and histopathological outcomes was made. Test accuracy was analysed using area under receiver operating characteristic curves reporting maximal sensitivity, specificity, and negative and positive predictive values. Task and domain specific analyses were explored including OCHRA and L-TMEpt identified events associated with inadequate outcomes.

Statistical analysis

Statistical analyses were performed using SPSS (v25.0, IBM, NY, USA) with Bland-Altman plots created using MedCalc (v18.6, MedCalc software, Ostend, Belgium) (184). L-TMEpt task areas, skill

domains and total score were handled as continuous variables with case complexity considered ordinal. Following descriptive analysis, internal consistency and inter-rater reliability for L-TMEpt total score was calculated with Cronbach's α and interclass correlation coefficient (two-way random effects model including 95% confidence intervals) respectively. No item weighting was implemented. Medians were compared with the Kruskal-Wallis test. Inter-rater case complexity reliability was calculated with cross-tabulation and Cohen's kappa coefficient (κ). As the L-TMEpt tasks and domains are not symmetrical, it was not considered appropriate to perform split-half reliability. For all analyses, three assessor and paired assessor comparisons were performed. Interpretation of correlation strength was quantified as excellent ≥0.75, good 0.74-0.6, fair 0.59-0.4 and poor ≤0.39. A p value of <0.05 was considered significant.

Data governance

A written data transfer contract between our group, the ALaCaRT trial management and Australasian gastrointestinal trials group were established. This facilitated anonymous trial data to be transferred and securely stored on a dedicated, encrypted, password protected research drive at the Clinical Research Unit, Yeovil District Hospital NHS Foundation Trust, UK. To comply with UK, European Union and Australian ethical and data governance considerations the unique trial ID was the only identifier transferred with creation of tables requiring patient identifiable data (such as demographic tables) performed by the clinicals trial unit, University of Sydney.

3.6 Video series and application of the laparoscopic TME performance tool

All videos held by the Sydney trials unit were supplied on portable electronic hard drives sent via secure international courier. All files had been previously deidentified and labelled solely with their unique ALaCaRT trial ID. All 115 that had been submitted to the central trial office, representing 53% of the 217 completed laparoscopic TME cases performed, were forwarded. Personal communication from the trial group confirmed that all videos held were forwarded and the number reflected a failure of recording and/or submission by the study sites and were considered protocol deviations.

During analysis, a further 16 videos were seen to be unsuitable as all contained significant omissions of TME performance preventing application of the L-TMEpt. Reflecting the study design and protocol, only seven cases included the abdominal phases of surgery. It was immediately evident there was considerable variation in the manufacturers and specification (screen size, standard or high definition) of the laparoscopic systems including the image capture technology across the trial centres. This corresponded to wide heterogeneity in the method of video capture with differing file types, lengths, quality and resolution resulting. All videos appeared unedited.

19 videos captured rectal transection and performance of the stapled anastomosis sufficient to allow completion of the L-TMEpt resection and anastomosis constituent task column. Therefore 99 ALaCaRT laparoscopic cases were limited to the first three L-TMEpt columns (all the TME dissection parts except resection and anastomosis) with 19 also assessed using the full instrument (table 3.4).

An independent assessor also performed blinded analysis of the same ALaCaRT video series after reviewing the L-TMEpt instruction manual and receiving training from Prof Hewett. Miss Sarah Abbott (Consultant Surgeon, Canterbury Health Board, Christchurch, New Zealand – assessor 2) reviewed the series during her senior laparoscopic colorectal fellowship under Prof Hewett at The Queen Elizabeth Hospital, Adelaide. No communication took place between the assessors who were blinded to all case, patient and outcome data. No data is available on her familiarity with assessment, the L-TMEpt or other clinical or research training prior to undertaking the video review.

9	121	206	269	347
13	125	**208**	272	351
16	128	209	273	363
22	**130**	212	275	367
26	**132**	213	279	370
30	**136**	216	280	**374**
33	137	219	281	379
70	**144**	222	283	384
74	146	223	290	386
77	147	226	**296**	388
80	150	**229**	307	392
87	153	**243**	**313**	394
92	157	244	316	**397**
98	161	**248**	318	403
99	162	252	321	405
106	167	**254**	327	**406**
107	184	256	331	408
110	190	261	332	409
113	200	263	340	416
115	203	267	344	

Table 3.4. The ALaCaRT case IDs for the 99 analysed videos. Bold numbers represent cases where the resection and anastomosis task phase were recorded allowing full tool use.

Revised statistical analysis plan

This information required amendment of the plan described in section 3.5. Dual analyses were undertaken. The three dissection task columns total (TMEdt) was calculated as this would allow uniform analysis of the 99 cases and allow direct comparison and combination with the cases reported in chapter 4. Given the L-TMEpt design, this approach would be expected to be valid for analyses of TME specimen quality as described in section 4.6.

Using the first three columns only (posterior, anterior and lateral mesorectal dissection) TMEdt scores between 12-48 are possible and where the whole tool was used 16-64. In view of this sample size, cases where full L-TMEpt completion was possible were combined with chapter 4 cases and presented in section 5.4 and appendix 3.

Individual interitem assessment was precluded as the other assessor's score data was provided as column totals and overall score alone without further detail. Internal consistency could therefore only be calculated for column totals and TMEdt. Through personal communication it was confirmed this data (or any other additional data) was never recorded and resulted from a misunderstanding of study requirements. Re-review of the videos was prevented by the associated extensive time requirements and competing clinical commitments.

3.7 Results

OCHRA training

Approximately 125 hours of OCHRA training and analysis was performed prior to independent analysis. My assessment of 20 previously captured laparoscopic TME cases showed excellent inter-rater reliability with previous OCHRA data (median 16 errors (range 7-34) vs. 17 (6-35), interclass correlation co-efficient 0.924 (95%CI 0.86-0.985), $p<0.001$) with no absolute difference >2. Test retest of three randomly selected cases confirmed reliability although the small numbers resulted in wide confidence intervals with associated impact on test p values (ICC 0.947 (95%CI 0.01-0.999), $p=0.05$; $r^s=0.667$, $p=0.5$). In a subjective summative assessment based on observed lap TME case review both trainers passed me as ready for independent laparoscopic TME OCHRA performance.

Analysis Time

The 99 videos contained a total of 325 hours of laparoscopic TME surgery (range 0.7-6). Where included, abdominal tasks were not reviewed as this was not required for L-TMEpt completion. Overall series analysis time was approximately 300 hours although completion of the L-TMEpt was quick (median 4 minutes, IQR 2-5). Test-retest reliability required a further 300 analysis hours.

OCHRA data

A total of 434 errors were identified during pelvic surgical phases of the 99 ALaCaRT cases (median 4, IQR 3-6, mode 3, range 0-11, Shapiro-Wilk 0.012).

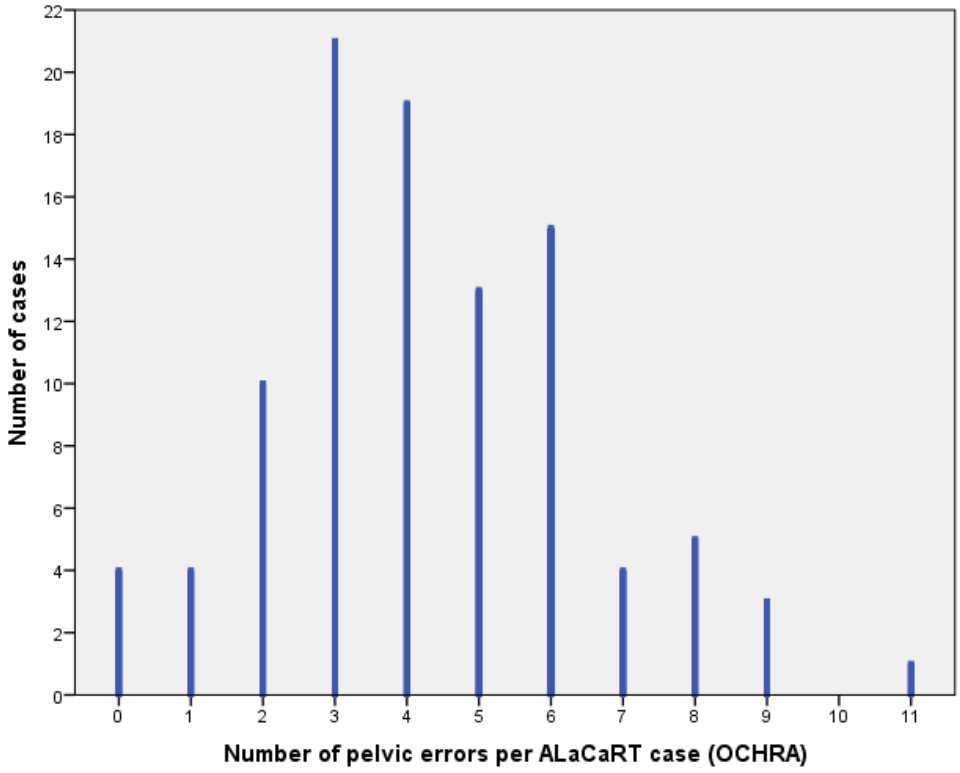


Figure 3.2: Bar graph displaying the distribution of OCHRA identified errors and case frequency.

	Sum (n)	%	Consequential (n)	Consequential %
Posterior TME	180	42.4	160	88.9
Anterior TME	87	20.5	75	86.2
Lateral TME	148	34.8	137	92.6
Resection and anastomosis	10	2.4	2	20
Total	425		374	88%

Table 3.5: Errors displayed by hierarchical task phase. Posterior and distal tasks accounted for the majority.

Errors were not equally distributed between the phases as defined by the hierarchical task analysis with most occurring during posterior and distal TME dissection ($p<0.001$). Few resection and anastomosis errors were encountered although this is highly likely to be secondary to the nature of the recordings that excluded the majority of these tasks (section 3.6).

The categorical breakdown of OCHRA data is displayed in tables 3.6 and 3.7. Dissection and instrument errors were more common that retraction and handling errors. Dissection in wrong tissue plane and in the wrong direction were the most frequent execution errors accounting for 23.5% and 12% of errors respectively. Too much force, instrument applied too close to structure and poor tip visualisation were also commonly seen. 88% of errors were consequential, primarily minor bleeding and tissue ooze (32.2% of consequential errors). Some 160 injuries to the mesorectum were observed although the majority were superficial with exposure or dissection into the rectal musculature rare events. It is noteworthy that 16 pelvic nerves injuries were observed with 23 risk to pelvic nerve events seen.

Regarding external error modes, three accounted for the clear majority of events: step done with too much force, speed, depth, distance, time or rotation (36.4%), step done in wrong orientation, direction or point in space (21.1%) and step done on/with wrong object (18.1%). 87.5% of error events were associated with the three commonly used surgical instruments for TME: ultrasonic dissector (58.4%), atraumatic grasper (17%) and hook diathermy (12.1%). 50% of pelvic errors occurred posteriorly with anterior (29%) and lateral locations less common (19.4%).

	Sum	%
Errors – Dissection/Instrument use		
Poor visualisation of tip	29	6.7
Overshoot of movement	19	4.4
Instrument applied with too little distance to structure	39	9.0
Inappropriate use of diathermy/energy source	15	3.5
Incorrect amount of energy applied	25	5.8
Dissection performed in wrong direction	52	12.0
Diathermy/dissection in wrong tissue plane	102	23.5
Use of inappropriate energy to dissect	12	2.8
Cutting without lifting tissues from underlying structures	10	2.3
Total	**303**	**69.8**
Avulsion of tissue	9	2.1
Too much blunt force applied to tissue	45	10.4
Traction applied with too much tension	22	5.1
Traction applied with too little tension	19	4.4
Traction applied in wrong direction	9	2.1
Inappropriate handling of tumour	2	0.5
Inappropriate grasping/blunt handling of structure	15	3.5
Use of inappropriate instrument to retract	10	2.3
Total	**131**	**30.2**
Combined total	**434**	**100**
Consequences		
Bleeding (ooze)	122	32.2
Bleeding (significant/pulsatile)	13	3.4
Mesorectal injury - breech of fascia only	73	19.3
Mesorectal injury – into mesorectal fat	74	19.5
Mesorectal injury – exposing rectal adventitia	6	1.6
Mesorectal injury – into rectal musculature	7	1.8
Rectal perforation	1	0.3
Diathermy burn to viscus	10	2.6
Sharp injury to viscus	0	0
Blunt bowel injury	5	1.3
Perforating bowel injury	0	0
Diathermy burn to other structure	0	0
Sharp injury to other structure	0	0
Risk of pelvic nerve injury	23	6.1
Injury to pelvic nerves	16	4.2
Injury to pelvic fascia	16	4.2
Injury to ureter	0	0

Consequences (continued)		
Risk of injury to other structure	8	2.1
Injury to other structure	3	0.8
Delay to progress of operation	0	0
Oncological compromise of operation	2	0.5
Total	**382**	**88**
External Error mode		
Step not done	11	2.6
Step partially completed	18	4.2
Step repeated	5	1.2
Second additional step	13	3.1
Second step performed instead	2	0.5
Step out of sequence	5	1.2
Step done with too much force, speed, depth, distance, time or rotation	155	36.4
Step done with too little force, speed, depth, distance, time or rotation	50	11.7
Step done in wrong orientation, direction or point in space	90	21.1
Step done on/with wrong object	77	18.1
Total	**426**	**100**
Instrument		
Hook Diathermy	51	12.1
Diathermy forceps	16	3.8
Ultrasonic dissector device	247	58.4
Johann grasper	72	17.0
Fine grasper	0	0
Swab	0	0
Suction	11	2.6
Scissors	4	0.9
Stapler	8	1.9
Bowel clamp	0	0
Clip applicator	1	0.2
Retractor	6	1.4
Other instruments	7	1.7
Total Instrument errors	**423**	**100**

Table 3.6: The OCHRA error, consequences, external error modes and instrument category lists used in all OCHRA studies throughout this thesis. The error data displayed refers to the analysed ALaCaRT cases.

Pelvic Error Location		
	Sum	%
Right anterior	60	14.1
Right lateral	35	8.2
Right posterior	48	11.2
Posterior	140	32.8
Left posterior	29	6.8
Left lateral	48	11.2
Left anterior	67	15.7
Total	**427**	**100**

Table 3.7: Pelvic spatial location of error events (figure 3.1).

Laparoscopic TME performance tool results

For all 99 cases, TMEdt mean score was 40 (83.3%, SD 4.64, range 30-48, possible scores 12-48, Shapiro-Wilk 0.076). Full tool assessment was possible in 19 cases (mean 57.7, 87.5%, SD 4.71, range 47-62, Shapiro-Wilk 0.171).

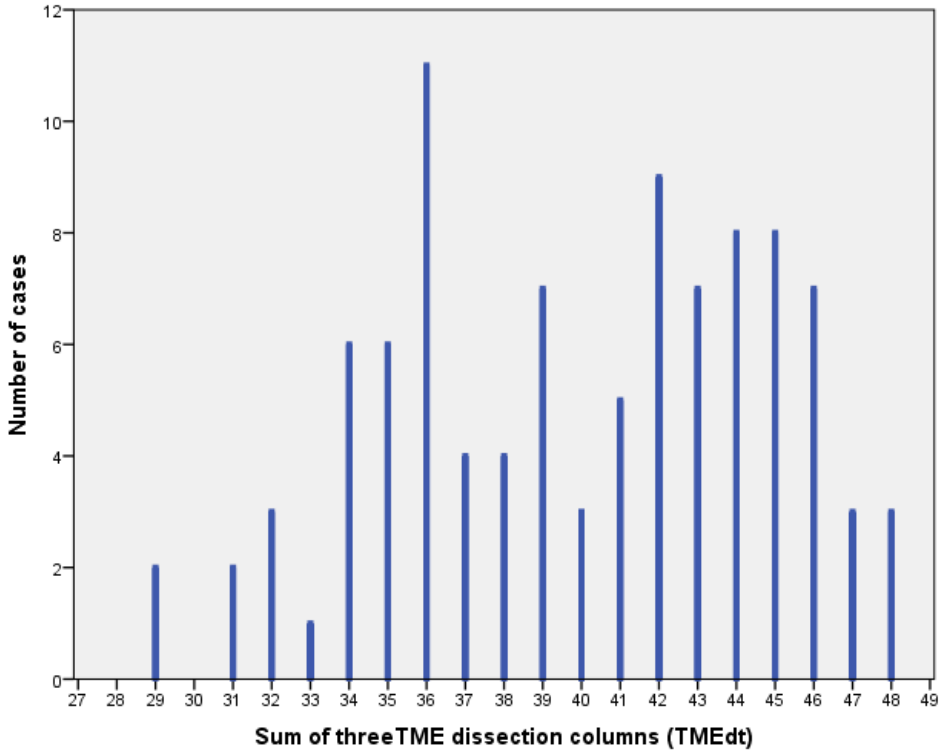


Figure 3.3: TMEdt score distribution for the 99 analysed ALaCaRT cases. Possible score range 12-48.

Mean (SD)	Posterior TME dissection	Anterior TME dissection	Lateral TME dissection	Resection and anastomosis	Total
N=	99	99	99	19	
Retraction + Exposure	3.42 (0.69)	3.22 (0.75)	3.33 (0.66)	3.57 (0.51)	14.57 (1.45)
Task performance	3.37 (0.63)	3.17 (0.73)	3.34 (0.66)	3.36 (0.84)	14 (1.71)
Errors	3.17 (0.85)	3.43 (0.77)	3.51 (0.61)	3.64 (0.75)	14.5 (1.51)
End-Product	3.4 (0.7)	3.09 (0.94)	3.47 (0.66)	3.64 (0.93)	14.64 (1.5)
Column Total	13.37 (1.78)	12.92 (2.08)	13.66 (1.83)	14.21 (1.58)	3 columns 39.95 (4.81) 4 columns 57.71 (4.03)

Table 3.8: Individual item, column, row and total score data. Figures represent mean and standard deviations. Consistently high scores are seen with no item averaging below three.

Observed scores were seen to be high. Of the 1188 total individual boxes completed, the mode individual item score was 4 with just 14 receiving a “1” (1.18%) and 146 (12.3%) a “2”.

Using the three point L-TMEpt complexity scale, the ALaCaRT case difficulties were not equally distributed. 26 cases were considered grade 1, 57 grade 2 and 16 as grade 3 ($p<0.001$). A significant decrease in TMEdt score was seen with each difficulty grade increase (median 44 (IQR 42-47) vs. 39 (36-42) vs. 36 (34-42), $p<0.001$, n=99).

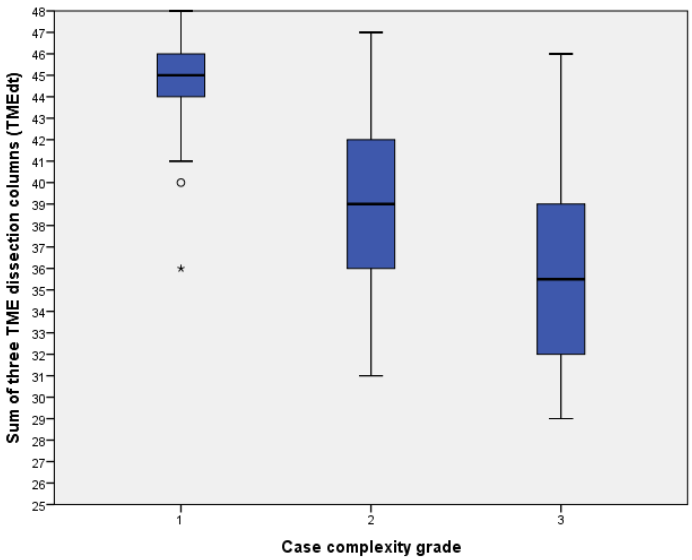


Figure 3.4: Box and whisker plot displaying TMEdt scores by case complexity grades. Boxes represent medians and interquartile ranges with whiskers depicting 98% confidence intervals. A significant drop in score is seen with each grade increase.

L-TMEpt inter-rater reliability

TMEdt score Posterior, anterior and lateral TME skill domains Possible range 12-48	Median	Interquartile range	Range
Assessor 1	41	36-44	25-48
Assessor 2	40	35-45	29-48

Table 3.9: The independent assessors returned comparable TMEdt scores.

The ALaCaRT TMEdt received high scores from both assessors with no evidence of significant inter-rater difference (Assessor one mean 40.3, SD ±4.75, assessor 2 40.1±5.17, p=0.821). Inter-rater correlation coefficient analysis showed good agreement for the TME dissection skill domains (ICC 0.598 (95%CI 0.42-0.73), p<0.001). Comparison of 13 complete L-TMEpt assessments by both assessors one and three showed excellent agreement (medians 57 and 59 respectively; ICC 0.77 (95%CI 0.25-0.93), p=0.008).

96 procedures received case complexity scores from both assessors with good agreement observed (κ=0.604, standard error 0.75, p<0.001). There was 77% absolute agreement with 22 results differing by one level.

TME Case Complexity Grade		Assessor 2			Total
		1	2	3	
Assessor 1	1	21	4	0	25
	2	10	43	2	55
	3	0	6	10	16
Total		31	53	12	96

Table 3.10: Inter-rater reliability for case complexity grading.

Inter-item reliability

The scores received from the second assessor were provided as task domain (column total) scores and L-TMEdt/L-TMEpt total. Therefore, inter-item reliability assessment was not possible. Table 3.11

displays assessor task domain scores with Cronbach's α. Analysis of the resection and anastomosis task area was not possible as only four cases had complete scores from all assessors.

	n	Assessor 1		Assessor 2		Cronbach's α (95%CI)	p
		Mean	SD	Mean	SD		
Posterior TME dissection	96	13.78	1.9	13.9	1.9	0.757 (0.625-0.837)	<0.001
Anterior TME dissection	95	13.1	1.9	13	1.9	0.689 (0.527-0.795)	<0.001
Lateral TME dissection	96	13.4	1.8	13.1	2.2	0.743 (0.610-0.830)	<0.001
Resection and anastomosis	19	14.3	1.1	14.5	1.2	0.506 (0-0.849)	0.118

Table 3.11: Internal consistency data for each L-TMEpt task domain.

L-TMEpt test-retest reliability

The 99 videos underwent blinded re-review an average of 12 (range 10-15) months after initial L-TMEpt assessment. Excellent positive correlation (r^s=0.796, r^2=0.614, p<0.001) and intraclass correlation coefficient 0.878 (95%CI 0.82-0.92, p<0.001) for TMEdt totals between assessments was seen. TMEdt score distributions were seen to be very similar between the assessments (both median 41, IQR 36-47, p=0.814).

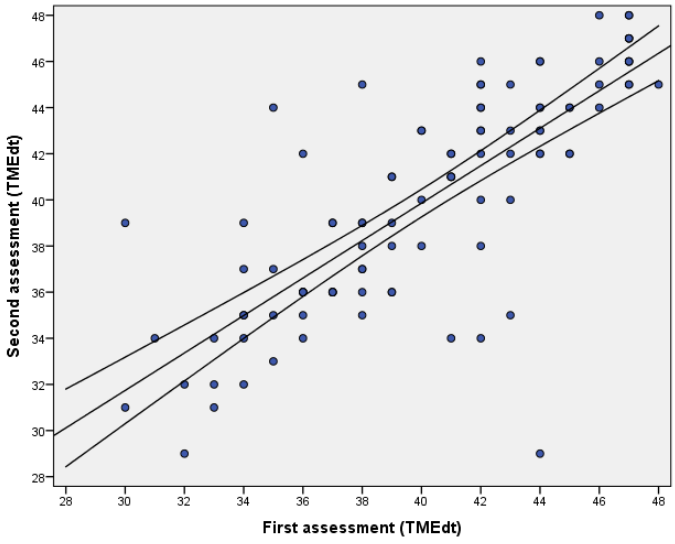
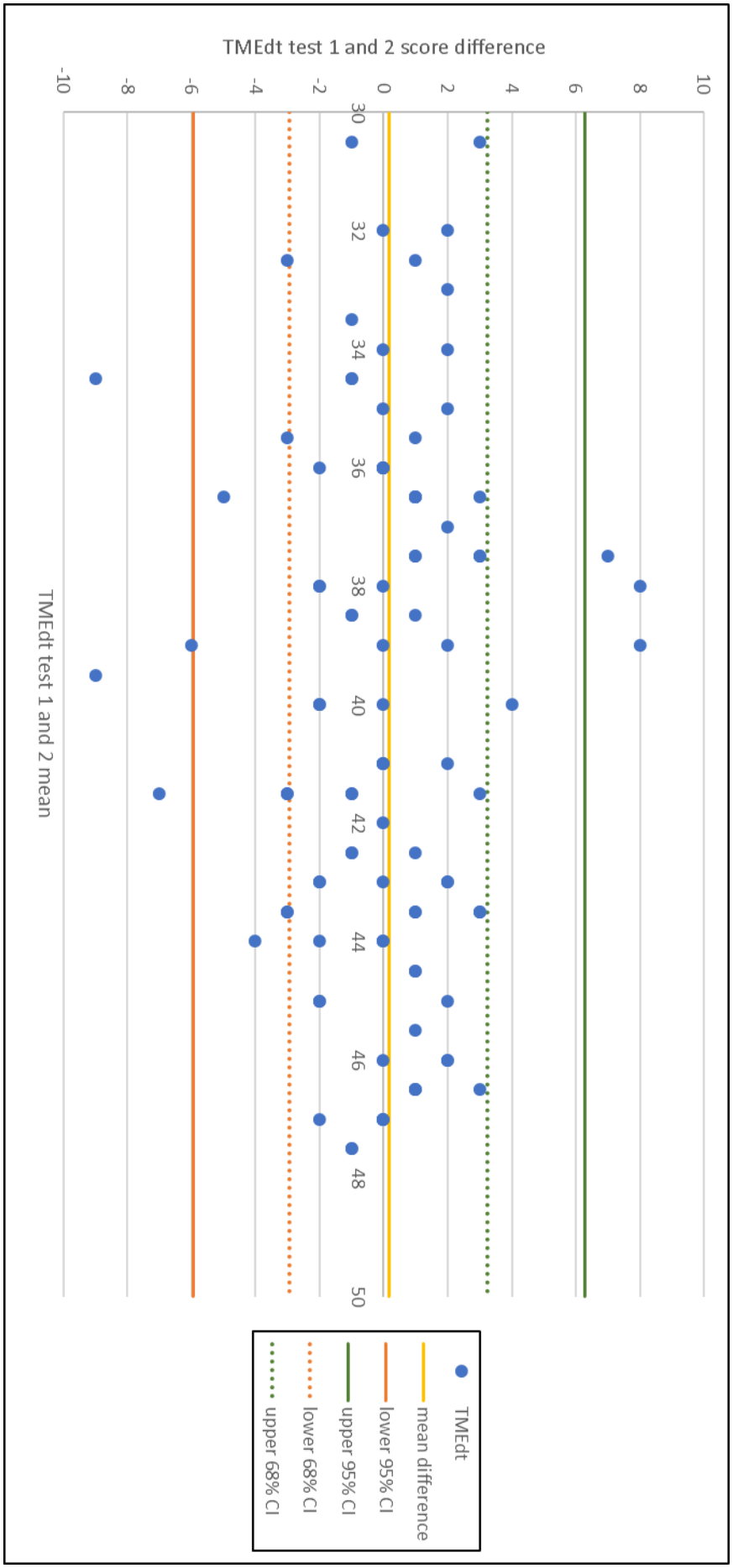


Figure 3.5: Scatterplot showing test-retest TMEdt scores with line of best fit and the 95% confidence interval. Excellent positive correlation is seen (r^s=0.796, y=7.38+0.81x, r^2=0.614, p<0.001).

The limits of this agreement were further explored through a Bland-Altman plot (184) (figure 3.6). The mean difference between assessments was 0.16 points (SD 3.12, 95%CI -5.95-6.27). A random scatter is seen with even distribution of points above and below zero and no funnelling effect suggesting no systematic difference exists between the pairs. The six outliers lying outside two standard deviations is in keeping with the sample size and 95% confidence interval presented. Only 10 cases fell outside one standard deviation.

Figure 3.6 (next page): Bland-Altman plot for the two TMEdt assessments. The mean difference was 0.16 points with the majority of differences falling within one standard deviation. The number of outliers fit with the 99 case sample size. The random distribution seen supports the test retest reliability of the tool.


TMEdt test 1 and 2 score difference
10
8
6
4
2
0
-2
-4
-6
-8
-10
30
32
34
36
38
40
42
44
46
48
50
TMEdt test 1 and 2 mean
TMEdt
mean difference
lower 95% CI
upper 95% CI
lower 68% CI
upper 68% CI

Comparing the 13 cases with two applications of the full L-TMEpt, moderate positive correlation with excellent intraclass correlation is seen (r^s=0.518, intraclass correlation coefficient 0.757 (95%CI 0.2-0.93), p=0.01) with similar scores recorded (1st median 57 (IQR 51-60), 2nd 59 (55-60)).

L-TMEpt case complexity test retest reliability

84 cases received identical case complexity grades with the 15 differing cases split equally between increasing and decreasing one grade. Good agreement was seen (Cohen's κ=0.738, p<0.001).

Number of cases		Case Complexity 2nd assessment			Total
		1	2	3	
Case Complexity 1st assessment	1	22	4	0	26
	2	4	49	3	56
	3	0	4	13	17
Total		26	57	16	99

Table 3.12: Test-retest reliability data for TME case complexity grading.

L-TMEpt inter-item reliability assessment

To allow comparisons with the inter-rater data, an identical approach was performed where task columns were analysed. Excellent reliability was observed for the TME dissection columns with the resection and anastomosis comparison limited by the available number of videos.

	Assessment 1		Assessment 2		Number of cases	Cronbach's α
	Mean	SD	Mean	SD		
Posterior TME dissection	13.8	1.8	13.4	1.8	99	0.844
Anterior TME dissection	13	2	12.9	2.1	99	0.772
Lateral TME dissection	13.4	1.8	13.7	1.8	99	0.880
Resection and anastomosis	14	1.2	14.3	1.6	13	0.341

Table 3.13: Test-retest internal consistency data for each task domain.

OCHRA and L-TMEpt concurrent validity

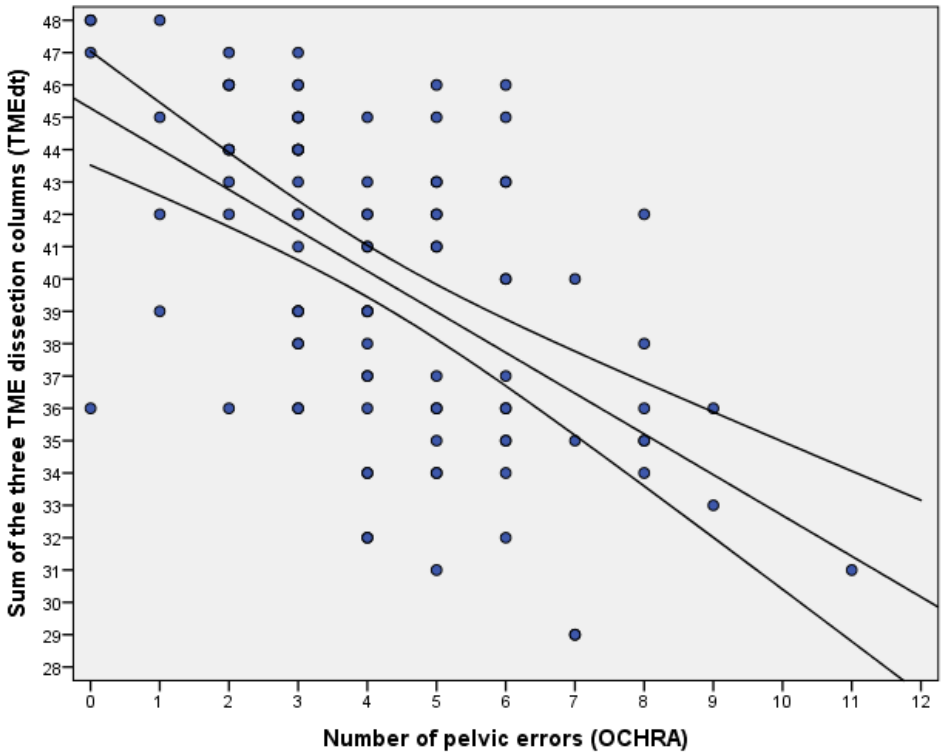


Figure 3.7: Scatterplot displaying OCHRA and TMEdt score data with line of best fit and 95% confidence interval. Moderate negative correlation is seen (r^s=-0.603, p<0.001). For clarity only a single dot is shown at points where there are multiple identical scores.

Moderate negative correlation was seen between the assessment tools (r^s=-0.603, p<0.001). Linear regression showed the y axis intercept at 45.3 with each additional error decreasing the L-TMEpt score by 1.33 (y=45.29-1.26*X, r^2=0.319). Given the narrow range seen in the number of pelvic errors, an additionally analysis was performed where this was managed as an ordinal variable. A similar result is observed with a significant six point drop in L-TMEdt score seen between cases containing three and four pelvic errors (p<0.001).

Pelvic Errors Total	TMEdt Median	N
0	47	4
1	41	4
2	43	10
3	44	21
4	38	19
5	39	13
6	38	15
7	38	4
8	38	5
9	39	3
11	30	1

Table 3.14: Tabular comparison of OCHRA pelvic errors and corresponding median TMEdt. The largest drop is seen between the three and four error point.

3.8 Clinical validity of objective assessment methods

78 (78.7%) cases met the ALaCaRT composite primary endpoint for a complete resection (table 3.2). The rate within the analysed sample did not significantly differ from the overall laparoscopic arm results (78.7% vs. 82%, p=0.781). Of the 21 (21.3%) that did not meet the endpoint, reasons included involved CRM (6, 6.1%), non mesorectal fascial plane of surgery (16, 16.2%) and involved distal margin (1, 1%). Two cases were histopathologically assessed as both CRM involvement and non-mesorectal fascial plane surgery (2%).

TMEdt scores did not differ according to tumour height (upper/middle/low, median 40 vs. 41. Vs. 40, p=0.945) or neoadjuvant therapy use (40 vs. 40, p=0.918). Higher scores were observed in women (42 vs. 40, p=0.024). There was weak negative correlation between BMI and L-TMEdt score (r^s=-0.405, p<0.001). No other demographic or tumour data was significantly associated with TMEdt score.

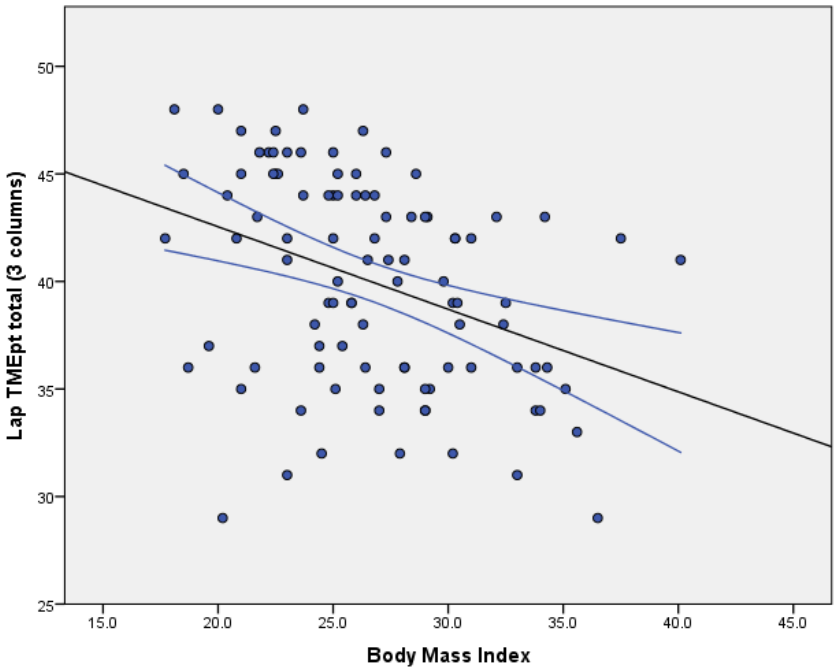


Figure 3.8: Scatterplot displaying TMEdt score and BMI with line of best fit and 95% confidence interval. Moderate negative correlation is seen (r^s=-0.405, p<0.001).

Histopathological composite endpoint

A significant six point difference was seen between cases with complete resection and those that did not meet all three components of the ALaCART primary endpoint (42 vs. 36, p=0.006, table 3.15). It was not possible to define a TMEdt score threshold with sufficient sensitivity and specificity for high stakes assessment in clinical settings (TMEdt AUROC 0.701, 95%CI 0.575-0.827, p=0.006).

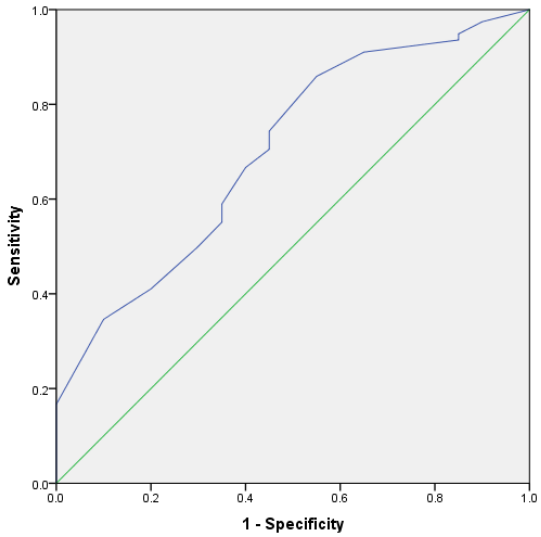


Figure 3.9: Area under the receiver operator curve for TMEdt score and successfully meeting the ALaCaRT composite primary endpoint (AUROC 0.701, p=0.006). TMEdt does not show sufficient clinical utility.

When the L-TMEpt was broken down into its constituent rows and columns, higher scores are seen in cases meeting the ALaCaRT composite endpoint with the exception of the underused resection and anastomosis column. All observed absolute differences were around one point with this reaching statistically significant in half of the eight columns and row comparisons. All other p values were <0.1 suggesting a type II error originating from the modest sample size. OCHRA pelvic errors were not different between the groups.

	ALaCaRT Composite Endpoint		
	Unsuccessful Median	Success Median	p
OCHRA Pelvic Errors	5	4	0.509
TMEdt	36	42	**0.006**
L-TMEpt	49	59	0.313
Sum Posterior TME column	13	14	0.083
Sum Anterior TME column	13	13	0.054
Sum Lateral TME column	14	15	**0.008**
Sum Resection Anastomosis	13	15	0.367
Sum Retraction Exposure Row	9	11	**0.015**
Sum Task Performance Execution Row	10	10	**0.014**
Sum Errors Row	10	11	0.085
Sum End Product Row	10	10	**0.037**

Table 3.15: L-TMEpt and OCHRA data displayed for the ALaCaRT primary endpoint.

The ALaCaRT composite endpoint was broken down into its three components. As there was only one involved distal margin data for CRM and mesorectal surgery is displayed in table 3.16.

	Circumferential margin			Mesorectal Fascia Plane		
	Clear	Involved		No	Yes	p
	Median	Median	p	Median	Median	
OCHRA Pelvic Errors (n)	4	5	0.77	5	4	0.403
TMEdt	41	37	0.265	36	42	**0.001**
L-TMEpt	59	(none)		49	59	0.105
Sum Posterior	14	14	0.982	12	14	**0.023**
Sum Anterior	13	14	0.616	12	13	**0.017**
Sum Lateral	14	13	0.197	13	15	**0.002**
Sum Resection Anastomosis	15	(none)		13	15	0.367
Sum Retraction Exposure	10	9	0.192	9	11	0.367
Sum Task Performance Exposure	10	10	0.286	9	10	**0.002**
Sum Errors	10	11	0.379	10	11	0.153
Sum End Product	10	10	0.364	9	10	**0.023**

Table 3.16: L-TMEpt and OCHRA data displayed for the CRM and mesorectal fascia plane specimens.

TMEdt scores are seen to significantly correlate with the plane of mesorectal surgery. Using the three TME columns, higher scores were seen for mesorectal fascial plane resections (42 vs. 36, p<0.001). A ten point difference was seen in the cases where full tool was possible although this did not reach significance again potentially representing a type II error. Five of the eight columns and row totals were significantly higher in mesorectal fascial plane cases with absolute differences of 1-2 points. Area under the receiver operator curve fell short of clinical utility requirements (AUROC 0.765, 95%CI 0.651-0.879, p<0.001).

The L-TMEpt did not identify CRM involvement with no significance score difference seen in any comparison. OCHRA case error totals were not associated with histological shortcomings with comparable counts seen between both assessments.

30 day morbidity

Operative assessment did not differentiate patients with uneventful recoveries and those that developed any 30-day morbidity event. No significant difference was seen in TMEdt scores (42 vs. 39, p=0.194) nor pelvic errors (4 vs. 4, p=0.605).

	Any 30 day morbidity?		
	None	Yes	
	Median	Median	p
OCHRA Pelvic Errors (n)	4	4	0.605
TMEdt	42	39	0.194
L-TMEpt	58	60	0.308
Sum Posterior	14	13	**0.028**
Sum Anterior	13	13	0.661
Sum Lateral	14	13	0.247
Sum Resection Anastomosis	13	15	0.331
Sum Retraction Exposure	11	10	0.076
Sum Task Performance Execution	10	10	0.522
Sum Errors	11	10	**0.032**
Sum End Product	10	10	0.345

Table 3.17: L-TMEpt and OCHRA data displayed for the 30 day morbidity endpoint.

No association was seen between TMEdt scores and the number of 30-day complications per patient (p=0.283) or highest Clavien-Dindo grade per patient (p=0.667). Statistically borderline TMEdt score differences were recorded in those that underwent emergency re-operation (35 vs. 41, n=6, p=0.055) or developed an anastomotic leak (35 vs. 41, n=3, p=0.086) suggesting a type II error. Weak inverse correlation between TMEdt scores and length of stay was seen (r^s=-0.353, p<0.001).

3.9 Discussion

Objective assessment of surgical performance was applied to a major randomised controlled trial as an intraoperative QA technique. This approach is required to meet the stated QA definition of a systematic process investigating whether the surgical care pathway meets quality criteria. Irrespective of the adopted quality standard, absence of intraoperative data represents a major omission with the associated loss of data that could identify potentially deficiencies and shape improvement strategies. Here intraoperative QA was successfully applied to a major TME RCT. This presents further opportunities including the expansion of intraoperative QA techniques to new areas including trial endpoints given their measurement of surgical performance and technique that would not routinely be captured by present methodological considerations.

Dual QA assessments were applied to 99 specialists performed laparoscopic TME case videos. This facilitated a developmental investigation into the reliability, concurrent and clinical validity of the L-TMEpt. Usability of the tool is shown and is entirely in keeping with early experiences during the L-TMEpt pilot and extensive use of the L-CAT within LapCo (74).

As described, both the number and nature of the trial videos presented unexpected issues meaning that alone they are considered inadequate to sufficiently address my objectives. Few videos contained resection and anastomosis tasks meaning the tool could not be applied as designed. There is a risk that score differences were artificially decreased with impacts upon statistical analysis and the validity of the study. However, use of the first three columns could be considered to remain a clinically valid experiment for TME specimen assessment as the dissection would have been completed at this stage and unlikely to be altered by the tasks assessed by the fourth column particularly as distal margin involvement was a very rare event in the trial (89). As shown in table 1.2 all laparoscopic vs. open TME trials, including ALaCART and ACOSOG Z6051, used histopathologically determined composite primary outcomes and so the TMEdt results remain of interest (40). Using an abbreviated version of the tool reduces the possible range of scores and may limit the ability of the

tool to differentiate between levels of performance. Although a significant difference was seen in a number of comparisons the absolute differences were small and their external validity is unclear.

Whilst overall tool scores were consistently high with little spread of the data, notable variation in scores was seen suggesting measurable differences in performance exist within this standardised procedure delivered by credentialed surgeons. This clustering of scores may impact the ability to differentiate between levels of performance in these specialist cases. Individual item scores of one and two were very low. It is unclear if this is a true representation of good surgery or the tool in its present form is inadequate with the need for revisions of grade boundary. It is possible that the tool is showing a ceiling effect which may limit use for specialist surgeon assessment (108, 126). The L-CAT was designed for a different level of surgeons and it is possible through adoption of similar methodologies the L-TMEpt maybe more applicable to senior trainees or consultants learning laparoscopic TME rather than independent, experienced surgeons.

A number of promising results are noted. TMEdt totals displayed histopathology validity with significantly higher scores seen in cases that met the ALaCART primary endpoint and particularly for mesorectal fascial plane surgery. This is a known strong predictor of oncological outcomes and the identified deficiencies in obtaining TME specimens resected along the mesorectal plane accounted for the main concerns of the ALaCART and ACOSOG Z6051 trials. This finding suggests, that subject to further dedicated study, quality assurance obtained via objective assessment of performance could be of use in specialist training, summative assessment and surgeon credentialing. The L-TMEpt showed less utility for morbidity or short term patient outcomes as no association was observed. Whilst the sample size risks type II errors there appear to be little difference in scores suggesting the tool may be best considered for intraoperative performance assessment alone which would be in keeping with its intended use and design. Importantly abdominal events are not captured.

Content based analysis of surgical video is a rapidly expanding field of study (185). The L-TMEpt scores here were awarded based upon unedited case video review. Within the research setting this

methodology allowed homogeneity and a cleaner dataset but it is unclear if scores would differ if the assessment was performed during or following observation, or participation in live surgery. OCHRA and L-TMEpt assessment were performed simultaneously meaning there is a risk they influenced each other. However, the excellent test-retest reliability data offsets this concern.

Comparison with OCHRA errors enacted during pelvic tasks showed moderate negative correlation with few outlying results and stronger correlation than that observed in for colonic procedures within LapCo (73). If the level of detail provided by OCHRA review is considered as the gold standard for intraoperative performance assessment, then this undertaking provides initial supportive evidence of L-TMEpt convergent and concurrent validity. The L-TMEpt has the considerable advantage of quicker completion times and could be performed at the end of a case avoiding time and labour intensive video review. During the extensive video analysis time I considered that purely counting errors represents an insufficient measure as they were seen to cover a broad spectrum of events and potential implications for the surgeon, procedure and potentially the post-operative progress of the patient. OCHRA does not capture presently capture these features which could be considered a limitation of the technique and one in need of remedy if application is to continue. Additionally, given the retrospective application of intraoperative QA it was not possible to influence surgeon behaviour such as technique or task order standardisation meaning analysis was limited to viewed events with the risk of procedural heterogeneity albeit reflecting real world experience.

Overall, this initial data supports the reliability of the L-TMEpt, but further study remains indicated. Test retest reliability showed excellent agreement for both scores and case complexity grading. Inter-rater reliability also returned supportive results with interclass correlations above the 0.7 threshold typically used for high stakes assessment settings. The second assessor was based in Australia and New Zealand which together with the wide international expert involvement in the design steps suggests the tool can be applied in external settings and healthcare systems. Training prior to L-TMEpt use may be an important consideration and it is not known if the other assessor

reviewed the instruction manual prior to tool use. This may also apply to the outlying examiner scores observed in the initial 12 case pilot work (129). No feedback data on the assessor's experience using the tool for these assessments is available representing an important omission.

The ALaCaRT study represents a well-designed, well performed, large scale, high quality surgical intervention RCT (89). It is seen to incorporate the majority of suggested QA methods for multicentre laparoscopic colorectal trials (110). The non-inferiority approach used is fully justified given the primary research question exploring whether laparoscopy represents a comparable technique to open TME (186, 187). The prior literature does not suggest superiority of a laparoscopic approach making a major superiority RCT with the associated long follow up and significant resource needs inappropriate. The application of QA has the potential to add insight to the trial results. This is explored in chapter 5.

In contrast to the CONSORT principles, the ALaCaRT publication does not provide any data on patients that did not enter the trial with only the number randomised stated (144). 475 patients, from 24 sites across 56 months equates to 0.35 patients per centre per month suggesting either non-consequential recruitment, large numbers not entering the study without explanation or low volume surgeons participating. A theoretical risk of selection bias results (155). There is a further question regarding whether Australasian data is directly applicable to UK patients and practice. Surgical training, MDT working practices and national guidelines are expected to be broadly comparable, but this has not been formally studied. Performing collaborative research between the UK and Australia presented many governance considerations. In addition to routine clinical research ethics, the handling of data required due consideration to the European Union general data protection regulations that came into force during this period. This resulted in a long delay between video assessment and receipt of the matched clinical data limiting subsequent follow up studies.

3.10 Conclusion

The application of objective intraoperative assessment of complex laparoscopic procedures within a major multicentre RCT was possible and seen to be a reliable quality assurance technique. Within this initial retrospective experience there is evidence that QA assessment scores are associated with clinically important post-operative outcomes. Review of the ALaCaRT videos series alone was insufficient to test my hypotheses and there remains an ongoing need to apply these methods.

The application of intraoperative QA techniques as endpoints within multicentre RCTs that contain interventions that may be subject to performance variation appears possible with potential advantages not offered by current trial methodologies. A dedicated feasibility study of this strategy is indicated.

Chapter 4 – Prospective application of intraoperative quality assurance within a laparoscopic total mesorectal excision trial

Intraoperative QA was successfully applied to RCT cases with initial supportive reliability and clinical validity evidence. The potential for prospective QA application within interventional RCTs was noted and could strengthen trial methodology and assist the interpretation of results. The feasibility of this approach is investigated with design and delivery of a bespoke multi-centre laparoscopic TME RCT was with intraoperative QA embedded from the start including as the primary endpoint.

Performing a multicentre trial is a major undertaking and was only possible with the dedicated help of a large team. I gratefully acknowledge the support and contributions from the following: Nicky Marks, Tressy Pitt-Kerby, Lucy Pippard, Linda Howard, Christy Felix, Jake Foster, Steve Gore (Yeovil District Hospital NHS Foundation Trust), Rowland Hackett (patient representative), Anne Bennett (Friends of Yeovil Hospital Charity), Prof Paul Ewings, Dr Amy Whitehead (senior statisticians, NIHR research design service), Emad Salib (Aid Medical), Liz Hawes, Ann Holmes, Sue Robertson, Gail Morrison, Joe Shoebridge, Sam Stefan (Portsmouth NHS Hospitals Trust), Daniel Jennings, Susan Sargent, Matt Dunstan, Claudia Forster (Royal Surrey County Hospital), Nicki Palmer, Leigh Davies, Matthew Williams, James Horwood, Buddug Rees (University Hospital of Wales), Richard Atkinson, Paul Lewis, Fred Dale, Lewis Thorpe, James Sturgess (Karl Storz Endoscopy UK Ltd). Stereopsis visual assessments were performed in conjunction with Morvern MacGillivray (Senior Orthoptist).

I am grateful to the participating surgeons for allowing their work to be scrutinised and despite extensive workloads provided their continual support: Mr John Conti (Portsmouth), Mr Iain Jourdan and Prof Tim Rockall (Guildford), Prof Jared Torkington and Mr Simon Phillips (Cardiff) and Mr Richard Dalton, Mr Andrew Allison, Mr Jonathan Ockrim and Prof Nader Francis (Yeovil).

This study was funded by a research grant from the European Association of Endoscopic Surgeons with delivery supported by UK Clinical Research Network portfolio adoption. In compliance with the grant contract, the trial manuscript was submitted to Surgical Endoscopy.

4.1 Copyright statement

This chapter contains a number of original tables and figures that were subsequently published as an open access article (188). They are presented here under the terms of the Creative Commons CC BY license (CC BY-NC-ND 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (188).

4.2 Introduction

In the previous chapter the intraoperative objective assessment tools were successfully applied to specialist performance of laparoscopic TME surgery with encouraging results. Intraoperative quality assurance has the potential to be embedded within surgical intervention trials but to date no dedicated studies have been performed. Such an approach could strength RCT methodology through assessment of procedural delivery including performance assessment and detection and measurement of any variation present. As no RCT has attempted this the feasibility is unknown.

Secondary to partial availability of the laparoscopic arm cases and heterogeneity within the analysed videos, the ALaCaRT TME series was considered insufficient to adequately explore my hypotheses. Prospective QA application to further homogenous cases together with matched clinical and histopathological data could strengthen the assessment tools validity and reliability explorations with the potential to support generalisability and external validity.

Aim

To prospectively apply intraoperative quality assurance as the primary endpoint in a laparoscopic TME RCT and evaluate the feasibility of this approach.

4.3 Laparoscopic TME case video searches

Attempts were made to obtain laparoscopic TME case video series. In addition to the data presented in sections 1.6 and 3.1, searches of www.clinicaltrials.gov, International Standard Randomised Controlled Trial Number registry (ISRCTN) and the UK clinical research network portfolio for the

terms “rectal” OR “rectal cancer” OR “total mesorectal excision” AND “laparoscopy” OR “laparoscopic” were performed on March 15th and 16th 2016. This loose search string was selected to minimise the chance of missing a potentially relevant study. No eligible study was identified.

Completed randomised controlled trials could be expected to hold comprehensive and robust clinical data. Additionally, multicentre studies would be typically larger with the additional need for quality assurance of intervention delivery (110). The 2014 Cochrane review on laparoscopic vs. open TME was studied (40). Of the 14 included RCTs, five were multi-centre studies (82-87, 189, 190). Only the COLOR II trial protocol described routine recording of laparoscopic performance (84, 85). In personal communication from the chief investigator (Prof. Jaap Bonjer, Amsterdam) it was reported that central review was rarely undertaken, and no video library was collated. All other multicentre RCT corresponding authors did not respond to email requests. The UK LapCo national training programme was designed for colonic cases and the Low Rectal Cancer Programme (LoREC) did not contain live operating meaning no video TME libraries were obtainable (71, 103). It was therefore concluded that a new bespoke laparoscopic TME study was required to further this research.

4.4 Conception and study planning

A multicentre approach was supported in keeping with the overarching aim of this thesis. This would be expected to increase the study size and generalisability of results. The safety concerns raised by the recent major laparoscopic vs. open TME RCTs all involved large number of sites suggesting this approach should be studied. There is the additional methodological need to quality assure minimal access surgical intervention delivery between centres (110).

Although a multi-centre prospective laparoscopic TME series would meet my needs, from the outset a randomised trial design was favoured. This would generate a higher level of evidence and reduce the theoretical risks of selection, performance, detection and attrition bias over a non-randomised design (155). The performance of randomised trials in the assessment of complex surgical interventions and technology efficacy evaluation should be considered the default option according

to the IDEAL collaboration (139). Undertaking an RCT was also favoured due to the need for comprehensive matched clinical and pathological data which would be most likely to be met through an RCT with the support of the available NHS research infrastructure including CRN portfolio adoption, trials unit support and specialist research design service.

Although justifiable from the educational research prospective alone, acquiring a laparoscopic TME series would represent a time and resource heavy undertaking which arguably outweighs the potential outputs. In this form, patients and surgeons may decline participation and further barriers to successful delivery such as lack of CRN portfolio adoption may be expected.

A laparoscopic vs. open RCT design was not appropriate as numerous large trials with subsequent long term follow up data and meta-analyses are already available (40, 91). Another trial arm, for example robotic, trans-anal or open surgery, even if a 2:1 or greater randomisation ratio favouring laparoscopy would also reduce the number of laparoscopic videos available for analysis. Therefore, suitable justification for an entirely laparoscopic TME multicentre RCT was sought.

4.5 Trial clinical research justification – 3D laparoscopy

An exciting opportunity was presented by an invitation from Prof. Alberto Arezzo (University of Torino) for my supervisors and myself to join the expert working group of the EAES consensus conference on 3D laparoscopy in surgery. This was initiated in response to the increasing use of 3D laparoscopic systems having reportedly overcome previously technical, resolution and user side effect issues seen in earlier platforms that limited clinical uptake (51, 191-193). The consensus consisted of a number of systematic reviews of the PubMed and EMBASE libraries to identify 2D vs. 3D RCTs and observational comparative studies. Included studies were assessed with the Cochrane risk of bias tool and the Grading of Recommendations Assessment, Development and Evaluation system (GRADE) (194-196). Statement and recommendations were drafted and once finalised subjected to voting at the EAES congress, London May 2018. As this work fell outside the aims of this thesis, full methodology and results are not presented here but available in the publication (197).

I contributed the training section where 72 studies containing 2452 participants who primarily performing box trainer tasks were identified and systematically reviewed. Task completion time and error rate were significantly better in 62% and 64% of studies allowing the evidence based statements *"the use of 3D imaging systems improves laparoscopic box trainer task performance task completion time and error rate but this benefit has not been studied in clinical practice (level of evidence – moderate)"* and *"3D vision improves outcomes for junior trainees performing standardized box trainer tasks using properly set up 3D high definition systems and passive polarized glasses (level of evidence – high)"* to be made (197).

This consensus conference together with a health technology assessment report defined areas for future research as well as establishing a number of areas of equipoise as no detrimental effects from contemporary 3D platforms was seen (193). Reflecting the variable quality of the available literature a number of limitations were identified and guidance for future research was provided *"there is a clear need for further randomised studies that use validated and reproducible tasks or standardization of intervention delivery. Wherever possible equipment, viewing distance, table height and ergonomics should also be standardised. Compliance with the CONSORT statement, A-priori sample size calculations, homogenous participant groups, surgical experience, stereopsis visual assessment, validated blinded assessment methods and robust randomisation tools are additional considerations that would strengthen results and add to our understanding* (197)*."*

Of direct relevance to laparoscopic TME and study planning, a number of clinical and experimental reports were included in the consensus that suggested stereoscopic platforms enhanced specialist performance with improvements more pronounced in complex task performance (198-200). The provision of depth perception could represent a route towards improved surgical performance with subsequent potential beneficial implications for patient outcomes. Any translation into improved TME specimen quality, reduction in morbidity, conversion or healthcare resources would represent an easily implemented and likely generalisable method for quality improvement. Intraoperative

analysis could also address the literature gap identified by the EAES consensus that *"no study attempted to assess quality of surgery* (197)*"*.

A single conference abstract of 40 3D anterior resections was identified pooled with 40 right hemicolectomies and 40 sigmoid colectomies performed by a single surgeon (201). They concluded operative time was not altered but no further detail or data was provided. The subsequent full text publication contained only their right hemicolectomy data (202). Therefore, there appeared to be no data on 3D laparoscopic TME surgery. Consequently, it could be considered that there is presently equipoise and that a 2D vs. 3D multicentre laparoscopic RCT could be clinically justified.

4.6 Methodology and study design

Strategy

There was no prior research on 3D laparoscopic TME to assist study design and in view of the available funding, timescale and laparoscopic systems, a definitive 3D TME RCT was not considered possible. In keeping with my aims a developmental RCT designed was adopted with aim of assessing the feasibility of embedding intraoperative QA into surgical RCTs. The study could also provide recruitment and outcome data to assist the design of future larger studies including effect size, sample size calculation, number of centres and expected recruitment metrics.

This strategy is compliant with the IDEAL recommendations for a phase 2b exploration study and the CONSORT principles were adhered to as well as recommendations for multicentre laparoscopic colorectal trials (110, 139, 144). Additionally, methodological recommendations for 3D laparoscopic studies were carefully reviewed and incorporated into the design and protocol (191, 192, 197).

The application of intraoperative QA as study endpoints

Arguably any impact of surgical technology is most likely to be observed in the intraoperative period. Investigation of the intraoperative period of 3D laparoscopic TME has not been previously performed and could represent a method for assessing the true impact of imaging technology on surgical performance and delivery intervention with new insight into results data.

In order to assess the feasibility of the application of intraoperative QA to RCTs OCHRA was chosen as the primary endpoint. The STARCAT RCT incorporated OCHRA assessment supporting this decision whereas the L-TMEpt requires investigation before wider application can be supported (129, 140).

A broad range of pre-defined secondary endpoints were captured as the optimal endpoint of a definitive 2D vs. 3D RCT is not known. They included operative factors (time, estimated blood loss, stoma creation and conversion – defined as inability to complete the dissection including the vascular ligation and/or requiring an incision larger than that needed for specimen extraction), surgeon cognitive load (203), histological specimen analysis (plane of mesorectal excision as previously defined in seminal publications (14, 204), lymph node yield, circumferential resection margin involvement and complete excision (R0)) and 30-day patient outcomes (morbidity, length of stay and unplanned readmission to hospital) (205) using the Clavien-Dindo classification (181, 183).

2D/3D trial hypothesis

Null: 3D laparoscopy does not alter intraoperative performance or the number of enacted errors during specialist performed TME surgery.

Alternative: 3D laparoscopy alters surgical performance and the number of enacted intraoperative errors during specialist performed TME surgery.

Sample size

From the methodological standpoint, feasibility and developmental studies do not require formal sample size calculations but should justify their chosen sample size (206). However, we experienced that peer-review of grant applications, trial registries, CRN portfolio eligibility assessors, ethics committees and reviewers for peer-reviewed publications, considered a sample size calculation to be a mandatory inclusion for consideration of support.

Therefore, to assess the impact of stereoscopic imaging on surgical performance the primary endpoint of this study was the total number of enacted intraoperative adverse events per case identified using OCHRA. As no 3D laparoscopic data was available, sample size calculation used a

combination of open and 2D laparoscopic TME cases taken from a completed Yeovil and Imperial College multicentre RCT that investigated the time from completion of neoadjuvant treatment on operative difficulty as well as data on the few rectal cancer cases present in LapCo (74, 140, 159). Both these studies performed OCHRA assessment to unedited case video. The average number of identified errors was 17 (standard deviation 7.02). Using a 5% significance level, a sample size of 62 had 80% power to detect a decrease in error counts to 12. This minimally clinically relevant difference of 30% difference was chosen based on observations in laparoscopic colectomy intraoperative performance within the UK LapCo national training programme sign off data. Allowing a 15% attrition rate for conversions or loss to follow up, the recruitment target was 72. This calculation was provided by Prof Paul Ewings (statistician and director NIHR research design service).

Timeline

To meet this target, using the UK national bowel cancer audit report to ascertain case volume it was estimated that four centres were required (9). One patient per site per month over 18 months would meet the target. Four sites were also felt an appropriate number for equipment provision, trial logistics and geographical considerations. A trial Gantt chart was produced (figure 4.1).

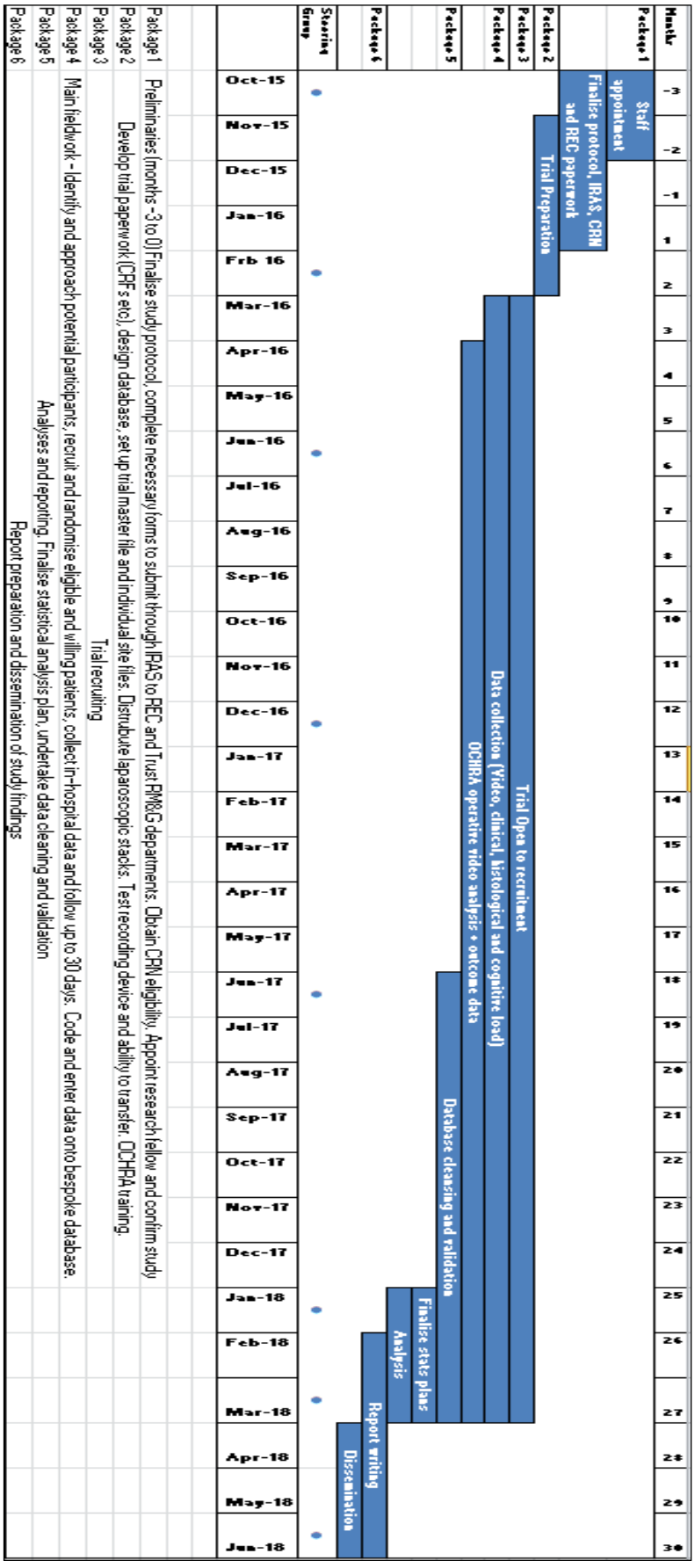


Figure 4.1: 2D/3D trial Gantt chart

Patient inclusion criteria

In order to successfully apply the intraoperative QA techniques and meet all required ethical and practicality considerations the following criteria were adopted: biopsy proven adenocarcinoma of the rectum (≤15cm from the anal verge), age ≥18, provision of written informed consent and the responsible local colorectal multi-disciplinary team advising elective laparoscopic TME undertaken with curative intent.

Neoadjuvant treatment use remained entirely at the discretion of the responsible clinicians. All patients were required to undergo minimum staging of pelvic MRI, staging CT chest, abdomen and pelvis, confirmatory tumour biopsy and full colonic assessment with either optical colonoscopy or CT colonography. Additional staging or peri-operative fitness investigations could be performed at the discretion of the responsible local clinicians.

Exclusion criteria were known or suspected inflammatory bowel disease, emergency, unplanned or palliative surgery, locally advanced cancers (T4a – TNM 5th edition), multi-visceral, pelvic exenteration or beyond TME procedure planned, refusal or inability to provide informed consent and concurrent or past abdominal or pelvic malignancy. With view to producing a homogenous procedural and video cohort with the ALaCaRT cases, abdominal-perineal excisions, trans-anal TME, robotic assisted cases and procedures where no anastomosis was planned were also excluded.

Surgeon inclusion criteria

Experienced, established high-volume UK laparoscopic TME centres were approached by the chief investigator for consideration of participation. Case volume was verified through the UK national bowel cancer audit project (9). Surgeons were required to have exceeded proficiency curve estimates and/or completed the UK LapCo training programme as participant or tutor (56, 71).

Surgeon stereopsis testing

Surgeon visual stereopsis assessment is an essential requirement for 3D laparoscopic studies (191). In keeping with standard practice of the Yeovil District Hospital ophthalmology department, surgeons took the TNO stereoscopic visual test (19th edition, Laméris Ootech BV, Utrecht, The

Netherlands). This consists of a series of plates with decreasing disparities from 480 seconds of arc down to 15 seconds. Participant stereo acuity was defined as the last correctly reported image with ≤120 seconds of arc considered normal.

Recruitment and randomisation protocol

In order to provide patients with sufficient time to consider participation and allow timely randomisation to assist theatre planning and equipment set up, approach and recruitment was encouraged once the patient was aware of their diagnosis and the recommendation for laparoscopic TME surgery. In line with requests made by patient and public participants, all approaches took place during planned hospital visits that formed part of routine pre-operative care pathways at each site.

To ensure allocation concealment, upon provision of consent, patients were centrally randomised. Allocation to the 2D or 3D arms used a pre-defined computer generated random integer generator (SPSS v24, IBM, Armand NY, USA). Given the sample size, no further stratification was undertaken.

Laparoscopic equipment set up

All trial cases were performed using Karl Storz IMAGE1 S D3-Link™ laparoscopic systems with zero and thirty-degree 10mm TIPCAM®1 SPIES 3D video laparoscopes. Images were displayed on 32-inch LCD HD screens (model EJ-MDA32E-K) and viewed with passive polarising glasses (Panasonic® Europe, Wiesbaden, Germany). To minimise cross-talk and facilitate optimal viewing and ergonomics, precise screen positioning and viewing distance was at the discretion of each surgical team. To maximise recruitment, generalisability of results and ethical acceptability to patients and surgeons, no constraints on patient selection, neoadjuvant treatment use, and timing of surgery was made. All perioperative patient care proceeded unaltered as per local policies. All centres reporting using established enhanced recovery after surgery protocols at the time of trial opening (207).

Intervention standardisation

All trial surgeons stated their usual operative plan matched the previously reported laparoscopic TME standardisation study of which a number were co-authors and/or expert participants (38). During study design we considered it unethical to place any restriction upon precise operative

technique, task order, instrument use or on table decisions which remaining entirely at the discretion of the operating surgeon.

Data Collection

Video recording utilised the advanced image and data acquisition system (AIDA™, Karl Storz Endoskopy GmBH, Tuttlingen, Germany) integrated into the laparoscopic stack. Entire cases were recorded unedited, deidentified (date, centre, any patient or surgeon detail removed) and labelled with the unique study ID generated at the time of randomisation as sole identifier. Irrespective of randomisation result all videos were recorded in 2D with sites instructed to avoid capturing any extra-corporeal views to ensure assessor blinding (194). Standard video output settings of the AIDA system were used (sequential 10 minute high definition MPEG4 files). Video files were securely transferred to the central trial office for analysis using portable hard drives (Canvio Basics, Toshiba Europe, Weybridge, UK). Upon receipt a second file coding was performed by an independent research administrator to further ensure blinded analysis which took place at least 90 days after the date of surgery. This pathway was agreed and tested with each surgeon and centre before opening.

Immediately following case completion, surgeons completed a series of 100mm visual analogue scales capturing overall case, task and pelvic complexity successfully applied to previous minimal access rectal cancer studies (140). The validated and commonly utilised NASA-TLX instrument was completed to assess cognitive load (203, 208, 209). Specimen analysis was performed at each site by specialist histopathologists blinded to trial arm and in keeping with UK Royal College of Pathologists reporting guidelines including the three-grade ordinal scale for plane of mesorectal dissection (14) (table 4.1). Patients were prospectively followed for 30 days by dedicated local CRN research staff independent of the trial management group and participating surgeons.

Plane of surgery	Description
Mesorectal fascia	The mesorectum should be smooth with no violation of the fascial covering. There should be a good bulk to the mesorectum both anteriorly and posteriorly and the distal margin should appear adequate with no coning near the tumour. Any defect should not be more than 5mm deep.
Intramesorectal	There should be a moderate bulk to the mesorectum with minor irregularity of the mesorectal surface. A moderate degree of coning of the specimen may be seen towards the distal margin. Importantly the muscularis propria should not be visible, except at the area of insertion of the levator muscles at the very distal aspect. There will be moderate irregularity of the CRM.
Muscularis Propria	There will be substantial areas where mesorectal tissue is missing with deep cuts and tears down onto the muscularis propria. On cross sectional slicing, the CRM will be very irregular and formed by the muscularis propria in places.

Table 4.1 Descriptions of the three grades of mesorectal dissection. Adapted from the dataset for histopathological reporting of colorectal cancer – Royal College of Pathologists (14).

OCHRA methodology

OCHRA was performed in keeping with the description provided in section 3.5. No attempt was made to capture or analyse extra-corporeal events such as patient set up, formation and closure of extraction site, resection of the specimen, placement of anvil device (to allow intra-corporeal stapled anastomosis) and where performed, creation of a defunctioning stoma.

Pelvimetry measurements

In order to assist validation of the three level L-TMEpt case complexity grade, pelvic dimensions and mesorectal area were calculated from pre-operative CT and MRI staging scans. Magnetic resonance imaging (MRI) was performed with a 1·5-T Magnetom machine (Siemens, Erlangen, Germany). T2-weighted images (T2WI) were obtained in high resolution transverse, coronal and sagittal planes; the transverse and coronal planes were angled exactly along the long axis of the tumour by using the sagittal plan. True axial T1 weighted images were also obtained. No rectal gel or contrast agents were used. Preoperative chest, abdomen and pelvis staging was performed using Toshiba 64-slice Aquilion CXL (Otawara, Japan) scanners capturing 0.8mm slices. Multiplanar reconstruction of

sagittal and coronal axes were made. All studies were viewed using the Carestream Vue Motion (v12, Carestream Health Inc, Rochester, NY, USA) or Sectra PACS (v18.2, Sectra AB, Linkoping, Sweden) digital platforms with all measurements obtained using the dedicated inbuilt software tools. Multi-planar reformatted axial and sagittal CT studies were simultaneously displayed to ensure appropriate landmarks were obtained based upon optimal and standardised orientations.

Three measurements were taken based upon the commonest measurement reported in the pelvimetry literature. A systematic review was performed using the PubMed library on December 19th, 2017 using the string "pelvimetry rectal". Studies were selected for full text review if they contained pelvic measurements and rectal cancer resectional surgery irrespective of study design, outcomes and surgical approach (210). Reference lists were further reviewed to identify other relevant studies. The number, nature and exact landmarks used were documented using a dedicated data capture form with measurements used by the majority of studies adopted. The identified studies and pelvimetry results are reported in the published manuscript (210).

Mesorectal area was calculated using a previously reported technique (211). A single T1 weighted slice at the level of the inferior point of the S5 body, perpendicular to the mesorectal tube was measured. In contrast to the previous report, the rectal luminal area was not measured as it could be considered irrelevant to the operating surgeon meaning no subtractions from the overall area were made (211). All scans were electronically transferred securely to the Yeovil research office and deidentified. All measurements were performed in triplicate. Linear and area figures were rounded to the nearest millimetre integer and one decimal place respectively. Inter-rater reliability was obtained by a second assessor with further independent validation performed by Dr Cen Thomas (consultant gastrointestinal radiologist, Yeovil District Hospital).

Pelvimetry measurement	Origin	Insertion	Orientation	Units	Scan
Pelvic Inlet	Sacral Promontory	Upper rim of symphysis pubis	Sagittal, midline	mm	CT
Interspinous	Distance between ischial spines		Axial	mm	CT
Mesorectal area	Single slice through mesorectum. Line originating at inferior border of S5, perpendicular to mesorectal orientation			cm^2	MRI

Table 4.2: Pelvimetry measurements applied including definitions and capture method.

Statistical analysis

In contrast to chapter three, L-TMEpt scores include assessment with all four columns of the L-TMEpt with possible scores between 16-64. Data was analysed using SPSS (v24.0; SPSS Inc, USA) and explored for normality with the Shapiro–Wilk test and detrended Q–Q plots and compared with parametric or non-parametric tests as appropriate. T-Test, Mann-Whitney U and Kruskal Wallis testing were used to compare medians from normal and non-normally distributed populations as appropriate. For categorical data, analysis included the use of cross tabulation, Fisher's exact test or chi-squared to test association between groups. Pearson and spearman rank correlation co-efficient were used to compare continuous and discrete ordinal variables. Data is displayed as medians with interquartile ranges unless specified. Analyses are reported as intention to treat except those solely based upon video analysis where the necessity for completed recorded surgical case required a per protocol approach. Kaplan-Meier curves were compared with the log rank method. Statistical significance was defined as $p<0.05$.

4.7 Regulatory and funding approvals

This trial was the first study to go through the new integrated regulatory approval service of the UK Health Research Authority following completion of their national rollout in spring 2016. HRA approval was dated 6th April 2016 (IRAS project ID 195969). The study was sponsored by Yeovil District Hospital NHS foundation Trust. The trial was prospectively registered (ISRCTN59485808).

Research ethic board approval was granted by the NHS South Central - Berkshire B Research Ethics Committee (reference 16/SC/0118) on February 24th, 2016. The study was approved for adoption to the UK NIHR CRN portfolio (cancer subgroup).

Study funding was provided by an EAES research committee grant awarded though open competition. All trial equipment was provided as a free loan for research from Karl Storz Endoscopy (UK) Ltd. Please see page 3 for full details and conflict of interest statement. Standard NHS indemnity contracts were signed with each site before receipt of laparoscopic equipment.

4.8 Trial delivery

Following site initiation visits in April and May 2016 the trial opened on 20th May 2016 when the green light was issued to Yeovil District Hospital. Queen Alexandra Hospital Portsmouth (3rd June), University Hospital of Wales Cardiff (26th August) and Royal Surrey County Hospital Guildford (5th January 2017) followed after local research and development capacity and capability approval, delivery of equipment, local theatre and research staff training and completion of delegation logs.

Substantial amendment

One month before the original planned trial closure (29th December 2017) an interim analysis of all endpoints in the then 77 recruited patients took place at the request of my supervisors and with the agreement of the sponsor. Attrition was 11.7% (9 patients) inside the inbuilt 15% margin.

No difference in the primary endpoint was observed (2D median 18 errors (IQR 13-21) vs. 3D 18 (12-22), p=0.703) however two of the underpowered patient centred secondary endpoints were seen to have clinically relevant differences that approached statistical significance.

		2D	3D	Total	p
Plane of TME surgery	Mesorectal Fascia	27 (77.1%)	31 (93.3%)	58 (85.3%)	0.085
	Intramesorectal	4	1	5	
	Muscularis Propria	4	1	5	
Total		35	33	68	

Table 4.3: Interim analysis for plane of TME surgery.

		2D	3D	Total	p
Readmitted to hospital	No	34	29	63	0.062
	Yes	3 (8.1%)	9 (23.7%)	12 (16%)	
Total		37	38	75	

Table 4.4: Interim readmission to hospital data.

Clinically important differences in the histopathologically assessed plane of TME surgery (mesorectal fascial plane 77% vs. 93%, p=0.085) and 30-day hospital readmission rate (8.1% vs. 23.7%, p=0.062) were evident but not statistically significant possibly representing a type II error. A short extension could significantly strengthen the output of the study as well as provide additional objective TME performance tool application and data to inform the design of a definitive 2D vs. 3D RCT.

In response to the steady recruitment and positive feedback from the chief investigator, my supervisors, CRN portfolio team, trial sponsor, participating surgeons, theatre teams and local research and development offices, a short extension was pursued. Available funding permitted a further three-month recruitment period. Written agreement including extension of contracts was kindly provided by Karl Storz UK Ltd. Request for a substantial amendment was made to the HRA on the 19th December 2017 and approved on January 10th, 2018 permitting recruitment until March 29th, 2018. The extension resulted in 11 further patients being enrolled.

4.9 Results

Recruiting surgeon and stereopsis testing

Ten consultant colorectal surgeons recruited to the trial (median 5.5 cases). Seven underwent stereopsis testing with no evidence of impaired stereo acuity seen (range 60-15 seconds of arc). Two others had previously obtained normal results on the Wirt Fly stereo test (Stereo Optical Inc®, Chicago, IL, USA) before contributing to a recent 2D vs. 3D laparoscopic cholecystectomy trial (212).

Recruitment and participant feedback

Steady recruitment was seen with a mean of four patients a month entering the trial (range 0-12). 88 patients were recruited in total. Subjective feedback from surgeons and centres was broadly positive

as trial participation did not impact on planned patient care, service delivery, their surgical intervention and all centres enjoyed provision of the 3D system. Just two patients declined participation when approached. One wished the surgeon to use their preference and one refused to consider participation in any study. Anecdotally a theatre team reported minor concern regarding the potential to forget to record the case however no serious adverse events were reported.

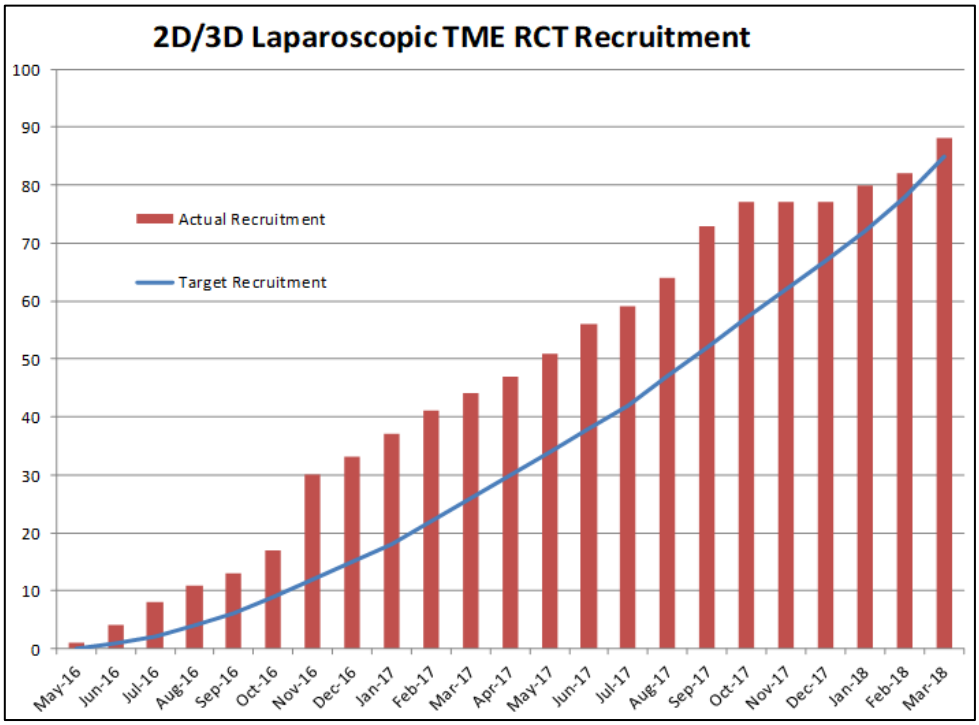


Figure 4.2 Trial recruitment graph. Steady and near continuous recruitment was achieved.

Trial CONSORT diagram and patient attrition

Performance assessment was possible in 77 (87.5%) patients that had a complete laparoscopic TME video produced through the trial falling within the planned 15% attrition rate. Reasons were in keeping with those expected with three (3.4%) not proceeding to surgery and six receiving different operations than originally planned (6.8%, 4 conversions to open, 1 abdominal-perineal excision of the rectum and 1 trans-anal TME). One theatre team forgot to start a recording and in one case the surgeon ignored randomisation and used 3D based on their subjective preference and additionally declined to record the case preventing analysis. In one case, the surgeon opted against performing

the planned anastomosis due to on table difficulties and concerns regarding adequate colonic blood supply although all other operative phases were recorded.

Patient screening logs were poorly completed at all centres limiting ability to assess numbers of potential patients, recruitment metrics and numbers of laparoscopic (and open) TME cases performed in each centre during the study timeframe.

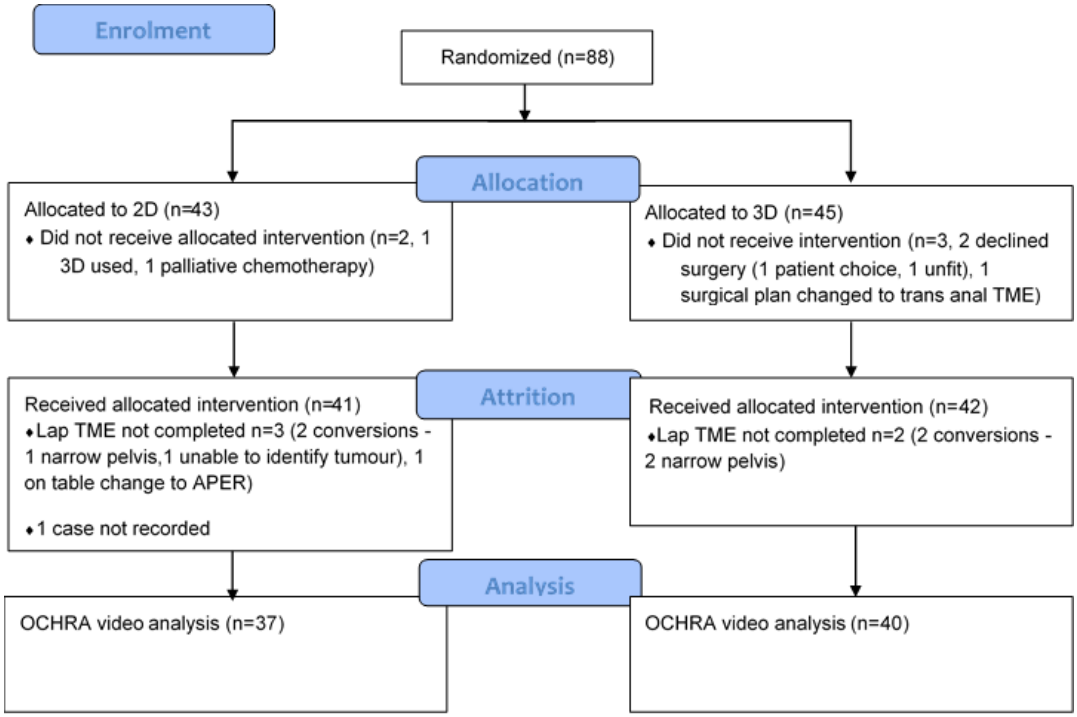


Figure 4.3 2D/3D trial CONSORT diagram

Patient and tumour demographics

Rectal cancer patient data is displayed in table 4.5. It is entirely comparable with UK practice as reported by the national bowel cancer audit project data (9). 20 (23%) underwent neoadjuvant treatment, all for threatened CRMs. The average tumour height (8.4cm) was slightly higher than the major laparoscopic TME trials probably reflecting the exclusion of abdominoperineal resection patients from this trial. There were no differences in any area between the 2D and 3D trial arms.

Table 4.5 (next page): Cohort demographic and tumour characteristics data for both arms. Cohort data is in keeping with UK practice expectations. There was no significant difference between any comparison (p values not shown). SD=standard deviation.

		2D			3D		
		Mean (sd)	Count	%	Mean (sd)	Count	%
Age		69 (11)			69 (10)		
Gender	Females		21	48.8%		16	35.6%
	Males		22	51.2%		29	64.4%
Body Mass Index (kg/m^2)		29 (5)			27 (4)		
Previous abdominal or pelvic surgery	No		29	67.4%		33	73.3%
	Yes		14	32.6%		12	26.7%
American Society of Anaesthesiologists physical status score	I		4	9.3%		2	4.4%
	II		24	55.8%		28	62.2%
	III		11	25.6%		14	31.1%
	IV		3	7%		0	0.0%
	Unknown		1	2.3%		1	2.2%
Neoadjuvant treatment	None		32	74.4%		36	80.0%
	Short course radiotherapy		1	2.3%		0	0.0%
	Long course chemoradiotherapy		10	23.3%		9	20.0%
Tumour height from the anal verge (cm)		8.5 (3)			8.4 (3.1)		
Tumour height from anal verge	Upper (10.1-15cm)		10	23.3%		14	31.1%
	Mid (6.1-10cm)		23	53.5%		18	40%
	Lower (≤6cm)		10	23.3%		13	28.9%
Predominant tumour location	Anterior		14	32.6%		11	24.4%
	Posterior		9	20.9%		7	15.6%
	Left lateral		8	18.6%		7	15.6%
	Right lateral		2	4.7%		7	15.6%
	Circumferential		9	20.9%		11	24.4%
	Unknown		1	2.3%		2	4.4%

4.10 Application of objective assessment tools to laparoscopic TME performance

OCHRA and the full L-TMEpt were successfully applied to 77 2D3D trial case videos. The data presented in this section refers to entire laparoscopic TME case analysis including all abdominal tasks as defined by the international standardisation report (74). No isolated pelvic analyses are displayed here in view of the points raised in this chapter's discussion and to avoid repetition with subsequent combined 2D3D and ALaCaRT analyses presented in section 5.3 and appendix 2+3.

OCHRA data

The total operative time within the trial was 419 hours equating to approximately 575 hours of video analysis. A total of 1377 error events were identified (median 18 per case, IQR 13.5-21.5, range 2-49). No differences in scores were seen across the 21-month analysis period.

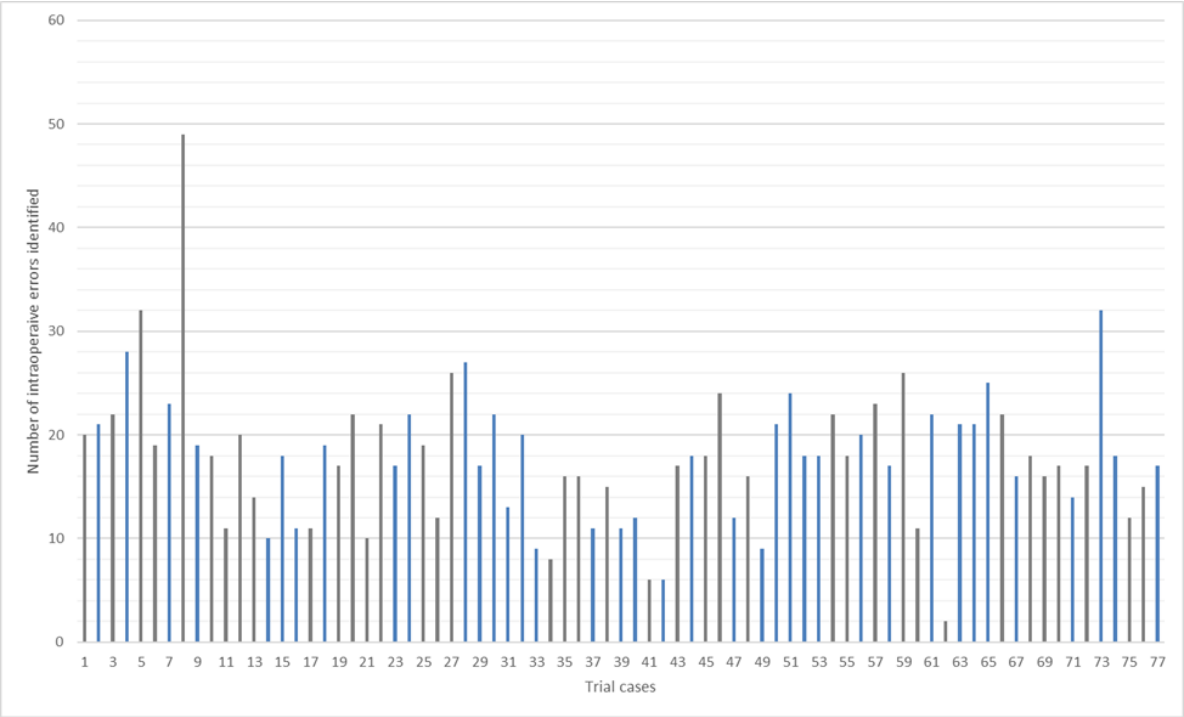


Figure 4.4: Sequential trial cases and OCHRA error count. Grey bars = 3D, blue bars = 2D.

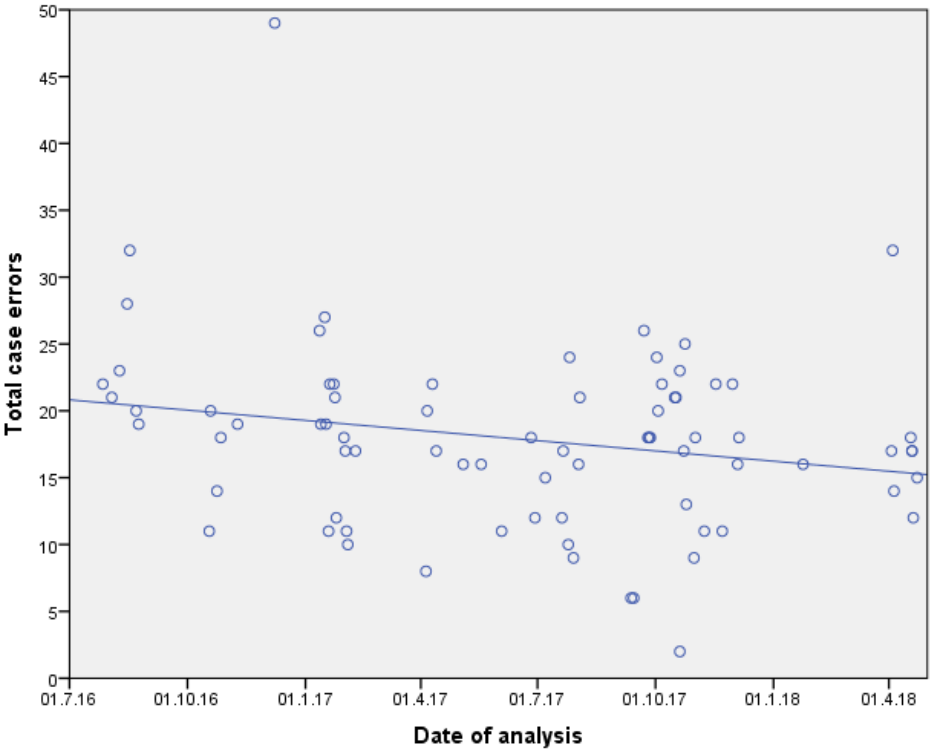


Figure 4.5: Scatterplot displaying error counts across the duration of the study with line of best fit. No pattern is seen suggesting consistent reporting with a near horizontal line of best fit (p=0.46).

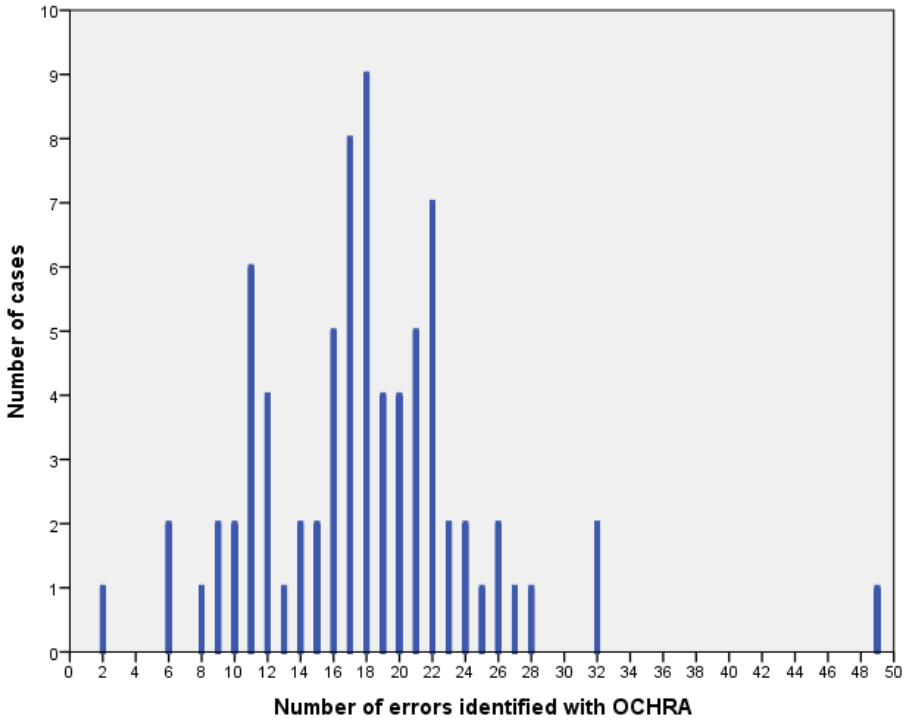


Figure 4.6: Bar graph displaying the number and frequency of OCHRA error counts per case

Errors were distributed equally between the pelvis (689 errors (50%), median 8, IQR 6-11, range 0-29) and abdominal tasks (688 (50%), median 8, IQR 5-11, range 0-18) with moderate-strong positive correlation between the number of abdominal and pelvic enacted errors per case (r^s=-0.688, p<0.001). Errors were distributed across all phases (mean 153 ± sd 77 errors per phase, p=0.2).

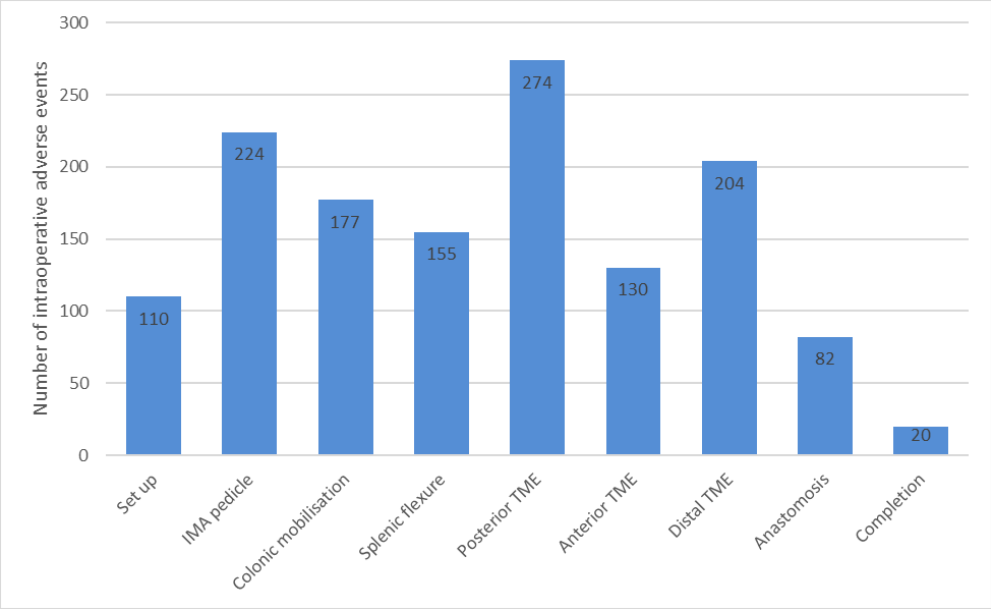


Figure 4.7: Location of identified intraoperative errors. There was an equal distribution between the abdominal and pelvic task phases.

OCHRA error analysis

The categorical breakdown of OCHRA data is displayed in table 4.6. Dissection and instrument errors were more common that retraction and handling errors. Dissection in wrong tissue plane was the most frequent execution error accounting for 20% of errors. Too much force, overshoot of movement and traction with too much force were also commonly seen. 78% of errors were associated with a consequence primarily minor bleeding and tissue ooze (n=462, 43% of consequential errors, 33.6% of all errors). 182 injuries to the mesorectum were observed although the majority were superficial with exposure or dissection into the rectal musculature were rare (n=18). It is noteworthy that in 25 cases (32.5%) an injury to pelvic nerves were observed with multiple injuries seen in six cases. There were ten errors (from nine cases) associated with potential oncological compromise of the operation.

Regarding external error modes, three accounted for the clear majority of events: step done with too much force, speed, depth, distance, time or rotation (36.6%), step done in wrong orientation, direction or point in space (24.7%) and step done on/with wrong object (16.3%). 90% of error events were associated with the three commonly used surgical instruments: ultrasonic dissector (39.3%; Harmonic ACE™, Ethicon Endosurgery, Johnson and Johnson Medical NV, Belgium and Thunderbeat, Olympus Europa SE & Co/ KG, Hamburg, Germany), Johan atraumatic grasper (30.4%) and hook diathermy (19.1%).

	Sum	
Errors – Dissection/Instrument use		
Poor visualisation of tip	91	6.6%
Overshoot of movement	112	8.1%
Instrument applied with too little distance to structure	108	7.8%
Inappropriate use of diathermy/energy source	31	2.3%
Incorrect amount of energy applied	91	6.6%
Dissection performed in wrong direction	68	4.9%
Diathermy/dissection in wrong tissue plane	281	20.4%
Use of inappropriate energy to dissect	46	3.3%
Cutting without lifting tissues from underlying structures	31	2.3%
Total	**859**	**62.4%**
Errors – Retraction/Tissue handling errors		
Avulsion of tissue	56	4.1%
Too much blunt force applied to tissue	161	11.7%
Traction applied with too much tension	112	8.1%
Traction applied with too little tension	40	2.9%
Traction applied in wrong direction	30	2.2%
Inappropriate handling of tumour	6	0.4%
Inappropriate grasping/blunt handling of structure	93	6.8%
Use of inappropriate instrument to retract	20	1.5%
Total	**518**	**37.6%**
Combined total	**1377**	

Consequences		
Bleeding (ooze)	462	43.1%
Bleeding (significant/pulsatile)	69	6.4%
Mesorectal injury - breech of fascia only	84	6.1%
Mesorectal injury – into mesorectal fat	80	5.8%
Mesorectal injury – exposing rectal adventitia	16	1.2%
Mesorectal injury – into rectal musculature	2	0.2%
Rectal perforation	6	0.6%
Diathermy burn to viscus	64	6%
Sharp injury to viscus	10	0.9%
Blunt bowel injury	30	2.2%
Perforating bowel injury	3	0.3%
Diathermy burn to other structure	22	2.1%
Sharp injury to other structure	4	0.4%
Risk of pelvic nerve injury	36	3.4%
Injury to pelvic nerves	35	3.3%
Injury to pelvic fascia	31	2.9%
Injury to ureter	0	

Risk of injury to other structure	45	4.2%
Injury to other structure	41	3.8%
Delay to progress of operation	23	2.1%
Oncological compromise of operation	10	0.9%
Total	**1073**	**77.9%**

External Error mode		
Step not done	47	3.4%
Step partially completed	66	4.8%
Step repeated	40	2.9%
Second additional step	24	1.8%
Second step performed instead	3	0.2%
Step out of sequence	8	0.6%
Step done with too much force, speed, depth, distance, time or rotation	500	36.6%
Step done with too little force, speed, depth, distance, time or rotation	119	8.7%
Step done in wrong orientation, direction or point in space	337	24.7%
Step done on/with wrong object	223	16.3%
Total	**1367**	

Instrument		
Hook Diathermy	261	19.1%
Finger switch diathermy	1	0.1%
Ultrasonic dissector device	536	39.3%
Johann grasper	415	30.4%
Fine grasper	9	0.7%
Swab	7	0.5%
Suction	17	1.2%
Scissors	17	1.2%
Stapler	39	2.9%
Bowel clamp	2	0.1%
Clip applicator	16	1.2%
Retractor	2	0.1%
Other instruments	41	3%
Total Instrument errors	**1363**	

Table 4.6: Whole case OCHRA categorical data from the 77 analysed 2D3D trial cases.

Laparoscopic TME performance tool data

77 videos were analysed with the full L-TMEpt. Completion of the tool was quick (median 5 minutes, range 3-8). Median total score was 52 (81.3%, IQR 47-55.5, range 34-64, possible score range 16-64). Similar to the OCHRA data, variation in scores indicate differences in technical performance.

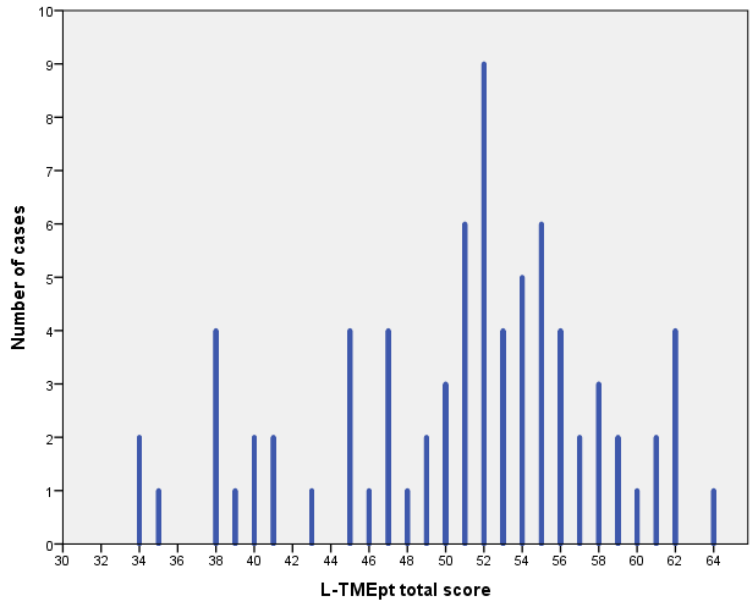


Figure 4.8: Bar graph displayed L-TMEpt scores for the 2D3D trial cases. Possible range 16-64.

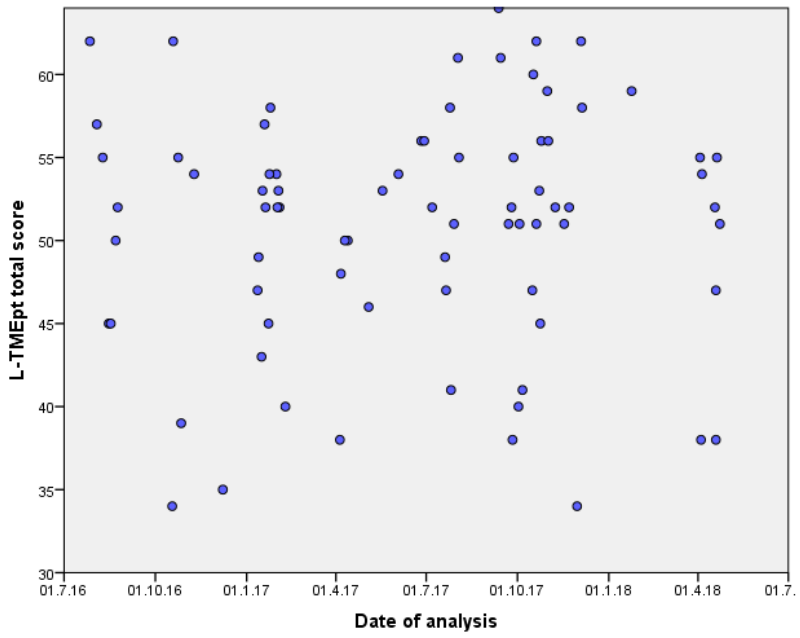


Figure 4.9: No difference in L-TMEpt scores was seen across the 21-month period (r^s=0.075, p=0.518).

Mean (SD)	Posterior TME	Anterior TME	Lateral TME	Resection and anastomosis	Row total
Retraction + Exposure	3.4(0.69)	3.27(0.62)	3.32(0.77)	3.39(0.86)	13.39(2.04)
Task performance	3.39(0.61)	2.95(0.78)	3.22(0.72)	2.84(1.16)	12.4(2.12)
Errors	3.03(0.78)	3.1(0.8)	3.23(0.79)	2.97(1.14)	12.34(2.03)
End-Product	3.35(0.68)	2.99(0.94)	3.51(0.68)	2.92(1.16)	12.77(2.41)
Column Total	13.17(2.12)	12.3(2.2)	13.29(2.19)	12.13(3.5)	**50.9(7.22)**

Table 4.7: Individual item average scores in the 77 2D3D trial cases. Overall, scores are comparable with the ALaCaRT trial, but five averages are shown to lie just under three.

Overall scores were again seen to be high and comparable with ALaCaRT data (TMEdt score (40 (36-42) vs. 41 (36-44), p=0.187)). The mean of 12 of the 16 individual items is seen to be above three. Standard deviations are wider than those observed in the ALaCART cases particularly in the resection and anastomosis column, possibly reflecting the smaller number of cases. There was excellent interitem reliability and internal consistency (Cronbach's α=0.914 (95%CI 0.884-0.94)).

1232 individual L-TMEpt items were graded in the 77 analysed 2D3D patients. The vast majority of scores were 3 and 4 (80.7%). 17 (1.37%) received a zero (not assessible) grade, all in the resection and anastomosis column of five patients in view of the small number of cases where an anastomosis was not performed. 23 items were scored with one (1.87%) with 198 (16.1%) scored as two.

Content validity – case complexity

Case complexity was not equally distributed in the trial and comparable to those seen in ALaCaRT with 46% scored as grade 1 (wide pelvis, no scarring or oedema in tissue planes, small tumour, mesorectum not bulky), 36% grade 2 (moderate width pelvis, minimal scarring or oedema, moderate tumour and or mesorectum bulk) and 18% grade 3 (narrow male pelvis, marked scarring and or oedema, bulky tumour and or mesorectum, p<0.001). Females received lower complexity grades with all grade 3 scores awarded to male patients (p<0.001).

Excellent inter-rater reliability was seen across all pelvimetry measurements (ICCs 0.852 to 0.953, 95%CIs range 0.568-0.995, all p≤0.001) (210). Case complexity significantly correlated with pelvimetry measurements. Male patients had significantly narrower pelvic dimensions although the absolute differences were small. The males also had significantly larger mesorectal areas.

	L-TMEpt case complexity						
Grade	1		2		3		p
N=	35 (45.5%)		28 (36.4%)		14 (18.2%)		**0.012**
Females	25		11		0		**<0.001**
Males	10		17		14		
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	
Pelvic Inlet (mm)	120	9	120	9	112	12	0.085
Interspinous Distance (mm)	107	16	103	12	92	10	**0.005**
Mesorectal Area (cm^2)	24.8	10.0	32.9	10.3	34.7	8.2	**0.002**

Table 4.8: Pelvimetry data displayed by case complexity grades. Harder cases contained more males and smaller pelvises.

	Females			Males			p
	Mean	SD	95%CI	Mean	SD	95%CI	
Pelvic Inlet (mm)	123	9	109-141	114	9	91-133	**<0.001**
Interspinous (mm)	114	11	92-143	93	9	72-107	**<0.001**
Mesorectal area (cm^2)	24.1	9.6	10.4-47.9	34.4	9.2	19.3-55.5	**<0.001**

Table 4.9: Pelvimetry data for both genders. Females are seen to have wider and taller pelvises with smaller mesorectums although the absolute differences were relatively small.

Case complexity showed a significant inverse association with total L-TMEpt score (p<0.001).

		L-TMEpt total score					
		n	Median	Interquartile range	Minimum	Maximum	p
L-TMEpt case complexity	1	35	55	52-58	39	64	**<0.001**
	2	28	52	45-45	38	62	
	3	14	47	38-50	34	58	

Table 4.10: Lower scores are seen with each increase in case difficulty grade.

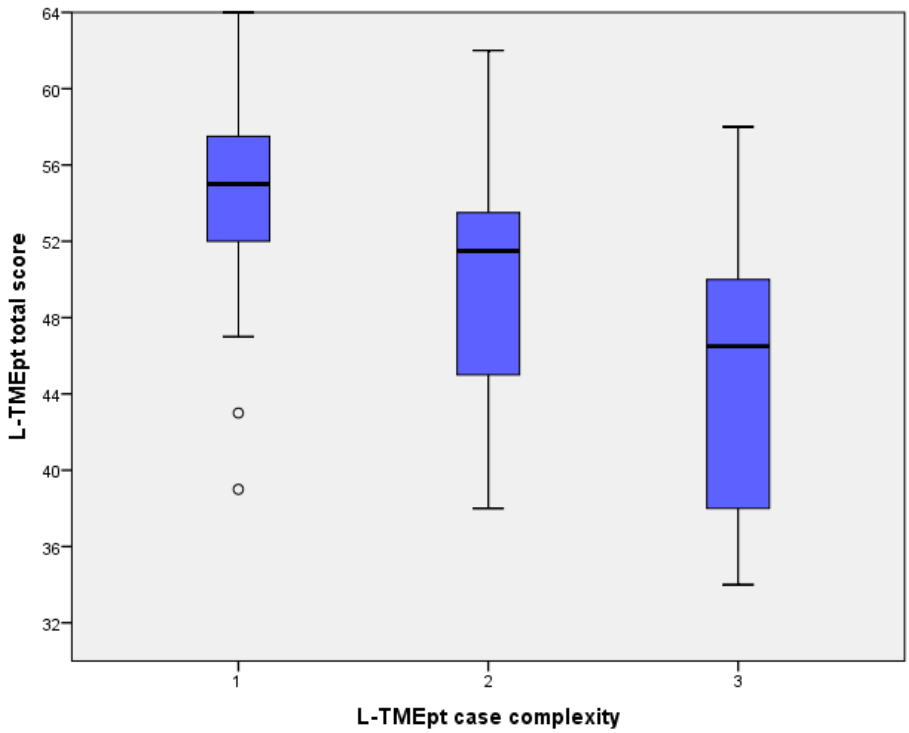


Figure 4.10: Box whisker plot depicting L-TMEpt scores for each case complexity grade. Boxes represent medians and interquartile range with whiskers covering the 98% confidence interval.

Validity of TME objective assessment

Given assessment was based on unrestricted live operating, face validity can be considered optimal.

Concurrent Validity

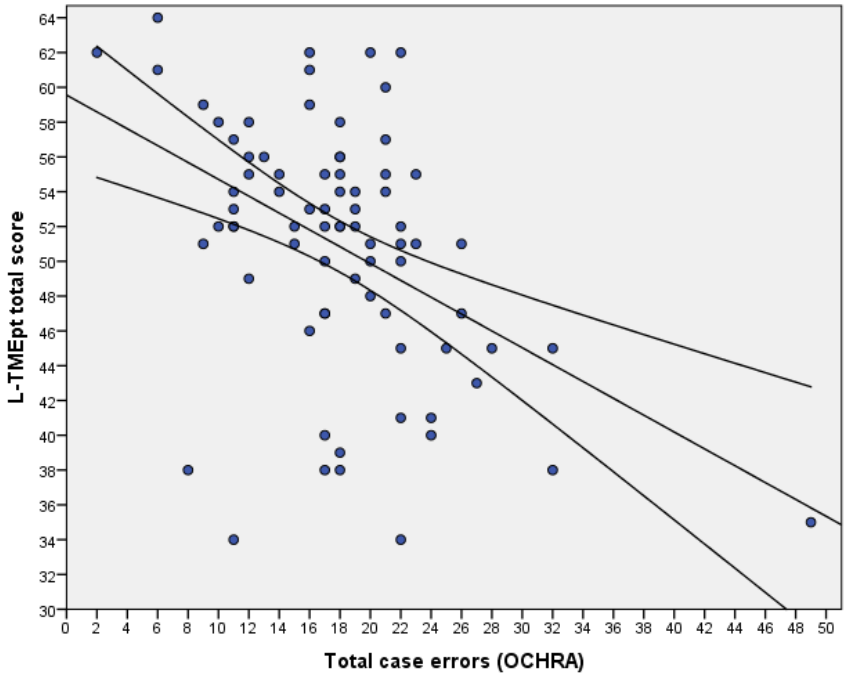


Figure 4.11: Scatterplot displaying OCHRA and L-TMEpt score data with line of best fit and 95% confidence interval. Moderate negative correlation is observed (r^s=-0.454).

Moderate negative correlation is seen between the two objective assessment methods (r^s=-0.454, p<0.001). Linear regression showed the y intercept at 59.6 with the L-TMEpt score decreasing by one for each additional two enacted OCHRA identified error events. A number of outlier cases are present with disparity in objective assessment score (goodness of fit, r^2=0.206). Variation in both objective performance assessment methods is again observed suggesting heterogeneity in technical performance. Consideration of the nature and severity of events may be required to assist the interpretation of this result.

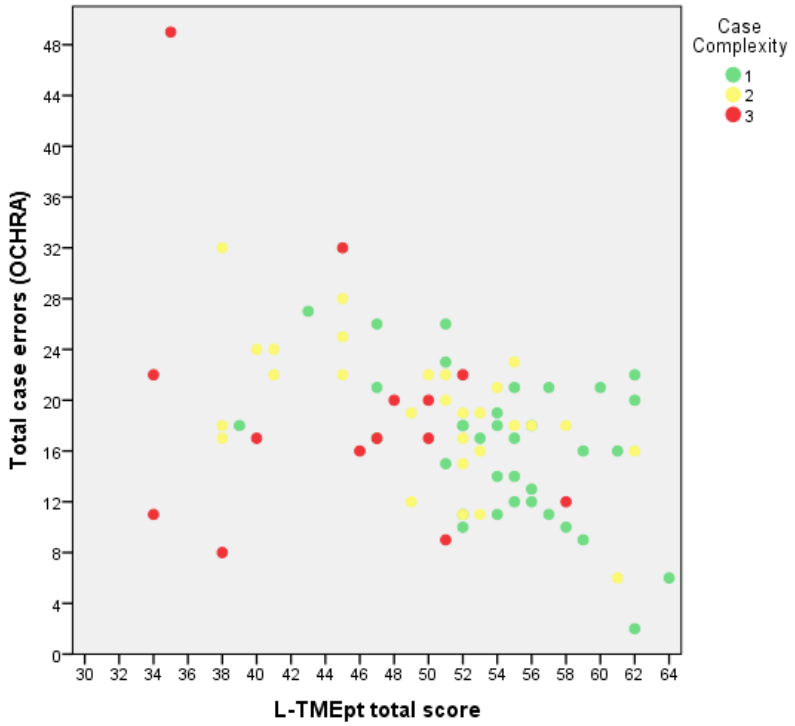


Figure 4.12: Scatter plot with L-TMEpt scores and OCHRA error counts with case complexity grades overlaid. A spread is seen across all three grades.

When case complexity grades are overlaid on combined OCHRA and L-TMEpt data, a wide and overlapping range of scores is seen across all three grades. This suggests grading does not represent the determining factor for objective assessment results. As both high and low scores were present across all three grades variable surgical performance might account for the observed differences.

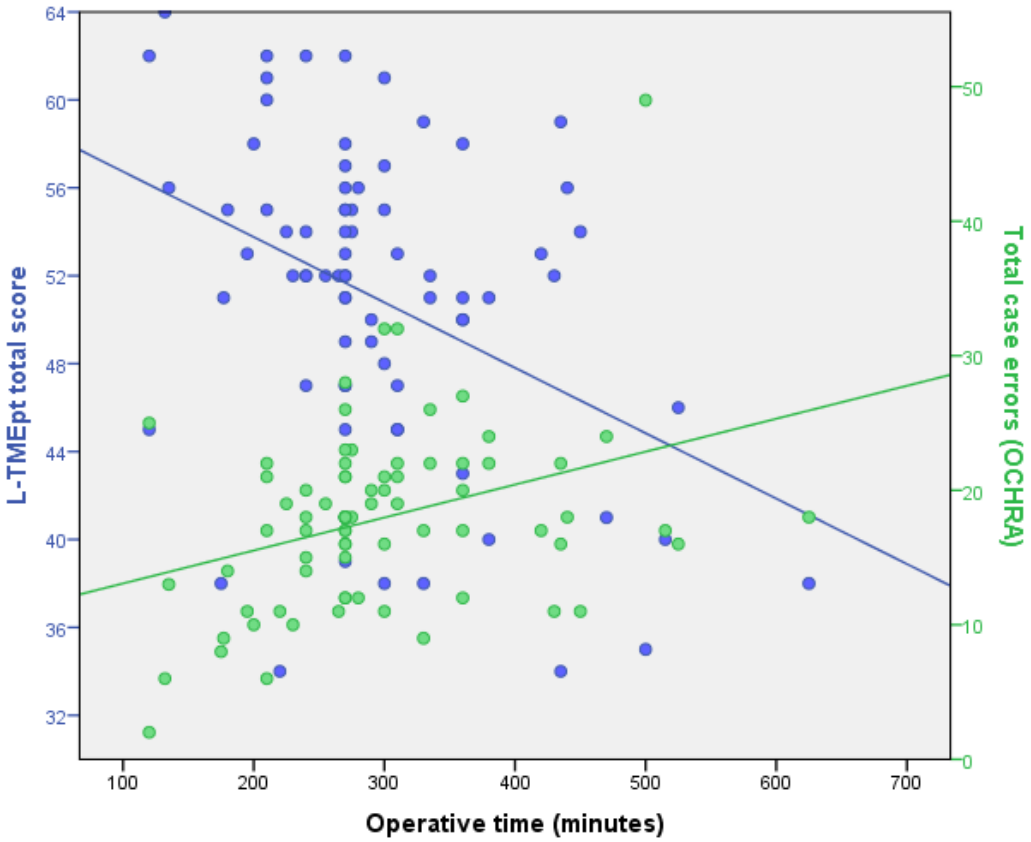


Figure 4.13: Dual axes scatterplot for L-TMEpt score (blue) and OCHRA errors (green) and operative time. Lines of best fit are shown but for clarity 95% confidence intervals are not displayed. As cases continue L-TMEpt score decreased and OCHRA error counts increased.

Weak positive correlation between operative time (median 270 minutes (IQR 240-335)) and total error counts is seen (r^s=0.35, p=0.02). Linear regression showed an increase of 0.02 errors per each additional minute of surgery with the y axis intercept point of 10.63 errors. Weak negative correlation between operative time and total L-TMEpt score is seen (r^s=-0.393, p<0.001) with a clustering around the 270 minute median with a wide range of L-TMEpt total scores observed at this point (39-62).

OCHRA error counts and L-TMEpt scores were not associated with mesorectal fascial plane surgery as reported by the specialist histopathologists (18 (10-19) vs. 18 (14-22), p=0.297 and 50.8 (±7.3) vs. 52.6 (±5.3), p=0.484 respectively).

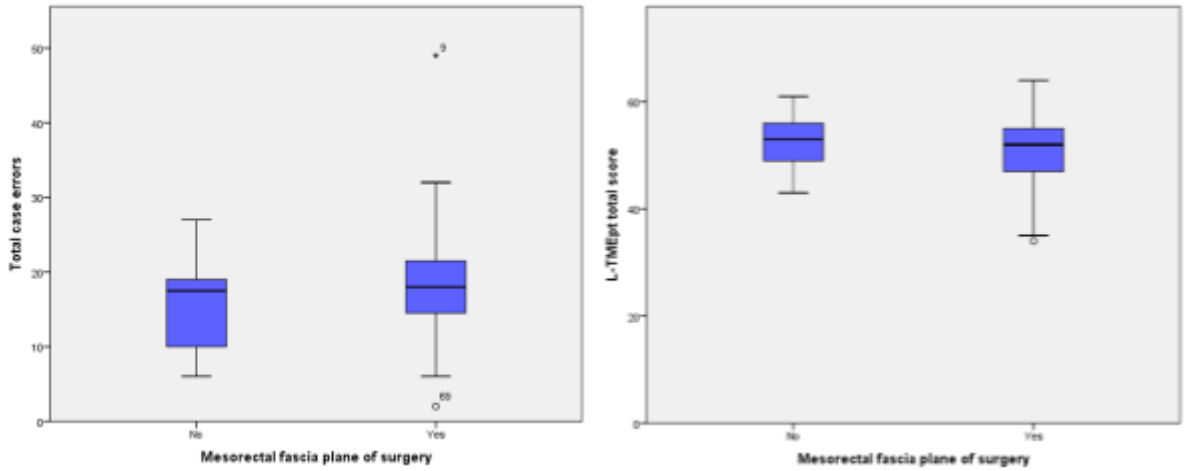


Figure 4.14: Box and whisker plots displaying mesorectal fascial plane specimens and a) OCHRA error counts and b) L-TMEpt score. No differences are seen between mesorectal fascial plane cases and intramesorectal or muscularis propria specimens.

N=77	TME Plane	Posterior TME		Anterior TME		Lateral TME		Resection and anastomosis		Row total	
		Mean (sd)	p	Mean (sd)	p	Mean (sd)	p	Mean (sd)	p	Mean (sd)	p
Retraction + Exposure	MF	3.42(0.7)	0.738	3.22(0.56)	0.259	3.37(0.95)	0.923	3.32(0.75)	0.652	13.3(2)	0.458
	Other	3.5(0.71)		3.4(0.84)		3.5(0.53)		3.4(0.84)		13.8(2.2)	
Task performance	MF	3.41(0.65)	0.82	2.98(0.75)	0.605	2.9(0.88)	0.686	3.15(0.76)	0.387	12.5(2.0)	0.904
	Other	3.4(0.84)		2.8(0.92)		2.92(1.2)		3.4(0.52)		12.5(2.1)	
Errors	MF	3.0(0.75)	0.649	3(0.81)	0.051	3(1.2)	0.355	3.15(0.83)	0.119	12.2(2.1)	0.274
	Other	2.9(1)		3.5(0.71)		2.9(0.74)		3.6(0.52)		12.9(1.1)	
End-Product	MF	3.41(0.65)	1	2.9(0.98)	0.237	2.95(1.2)	0.459	3.56(0.65)	0.786	12.8(2.4)	0.756
	Other	3.4(0.7)		3.3(0.82)		3.2(1.14)		3.5(0.71)		13.4(1.6)	
Column Total	MF	13.3(2)	0.849	12.1(2.2)	0.222	12.2(3.7)	0.661	13.2(2.2)	0.427	**50.8(7.3)**	0.484
	Other	13.2(2.6)		13(1.83)		12.5(2.3)		13.9(1.7)		**52.6(5.3)**	

Table 4.11: L-TMEpt means and standard deviations for mesorectal fascial plane specimens and other cases. No difference is seen in any comparison.

There was no difference in the total L-TMEpt score nor any individual, task area or skill domain score between TME specimen judged to be in the mesorectal fascial plane and intramesorectal or muscularis propria surgery.

Using OCHRA (both total errors and pelvic errors) and L-TMEpt data to define a clinically applicable threshold was not possible given the variation and wide distribution of non-mesorectal fascial plane case scores. Receiver operating characteristic curve showed discriminatory thresholds could not be established for mesorectal fascial plane surgery using L-TMEpt score (AUROC 0.431 (95%CI 0.258-0.603), p=0.485) and both OCHRA and pelvic error counts all failing to show clinical utility (total case errors AUROC 0.603 (95%CI 0.49-0.807), p=0.298 and pelvic errors AUROC 0.55 (95%CI 0.349-0.751), p=0.615). This either shows performance variation within the trial or the QA methods are not valid.

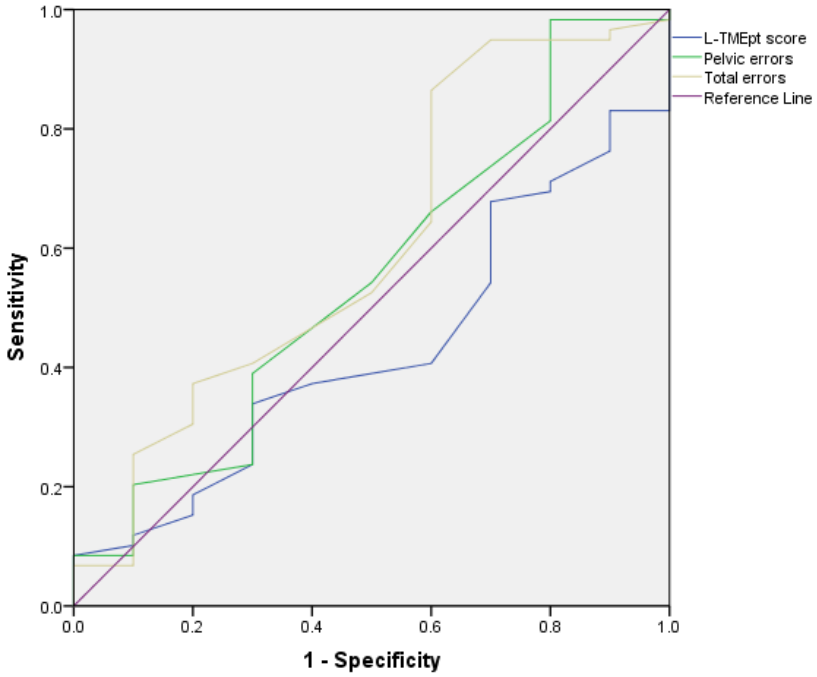


Figure 4.15: Area under the receiver operator curve for the objective assessment methods applied. None show sufficient discrimination to be considered for ongoing clinical or research use.

30 day morbidity

In the 77 cases a total of 99 morbidity events occurred within 30 days of surgery.

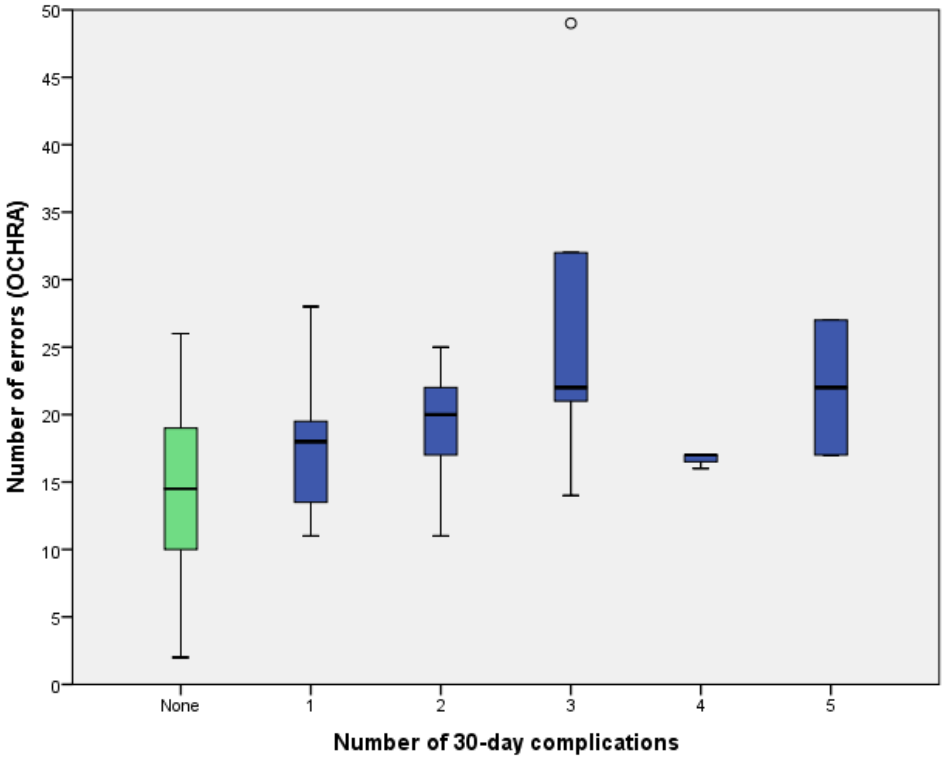


Figure 4.16: Box and whisker plot for OCHRA error count for patients grouped by the number of 30-day complications developed. (patient numbers per group were: none 34, 1 17, 2 18, 3 10, 4 4, 5 2)

There was a significantly lower number of case errors seen in patients that developed no complications (15 (10-19) vs. 19 (17-22), p=0.002) with more errors in patients that developed higher numbers of complications (p=0.007). There was no difference in total case error counts in those that developed an anastomotic leak (median 17 vs. 18, p=0.651) or required re-operation (17 vs. 18, p=0.782). Total case errors showed weak positive correlation with post-operative length of inpatient stay (r^s=0.339, p=0.003). An increase in total case error count were seen in patients were grouped by the highest observed Clavien-Dindo event (p=0.026).

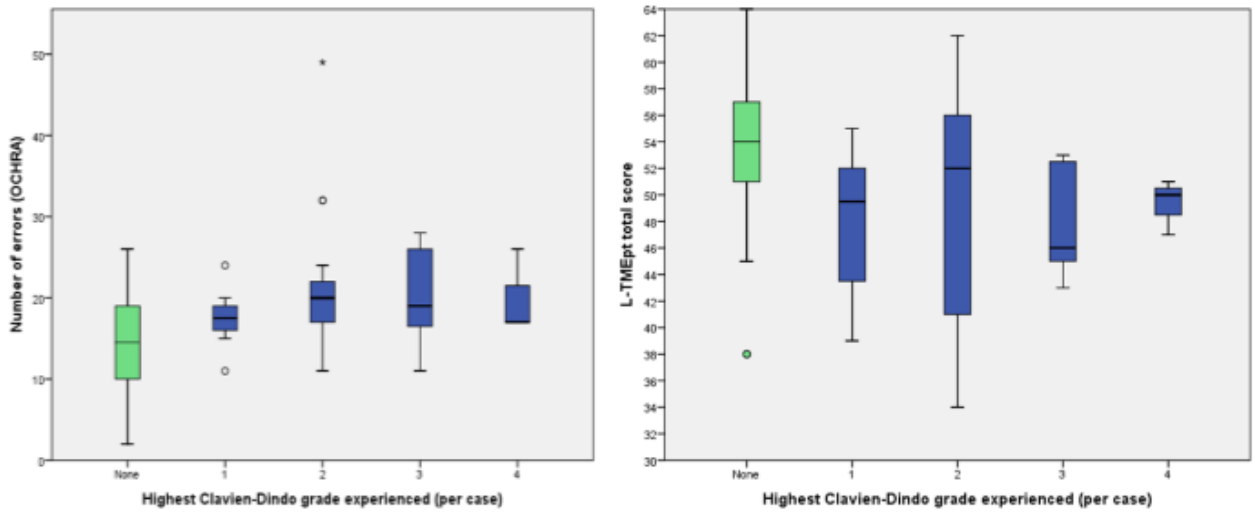


Figure 4.17: Box and whisker plot with patients grouped by the highest Clavien-Dindo grade developed within 30 days of surgery a) OCHRA error count and b) L-TMEpt score. (n=None 34, I 9, II 32, III 7, IV 3). Boxes represent median and interquartile range with whiskers covering the 98% confidence interval.

A small but significantly higher total L-TMEpt was seen in cases with no complications (median 54 (IQR 51-57, range 38-64) vs. 51 (45-54, 34-62), p=0.015). A small decrease in L-TMEpt scores were seen with each additional complication recorded although this was of borderline statistical significance (p=0.055). A non-significant decrease in scores were seen when patients were group by their highest observed Clavien-Dindo event (p=0.08). No difference in scores were seen in patients that leaked (47 (46-52) vs. 52 no leak (45-53), p=0.242) or underwent re-operation (49 (45-53) vs. 52 (47-56), p=0.352). The was no correlation between length of stay and L-TMEpt score (r^s=-0.108, p=0.349). There was a small but statistically significant drop in scores seen in patients readmitted to hospital (51 (41-52) vs. 53 (48-56), p=0.022).

	Morbidity	Posterior TME		Anterior TME		Lateral TME		Resection and anastomosis		Row total	
		Mean (sd)	p	Mean (sd)	p	Mean (sd)	p	Mean (sd)	p	Mean (sd)	p
Retraction + Exposure	None	3.6(0.62)	**0.039**	3.3(0.7)	0.621	3.5(0.68)	0.106	3.23(0.94)	0.143	13.63(1.92)	0.456
	Any	3.28(0.71)		3.26(0.57)		3.21(0.81)		3.49(0.8)		13.23(2.12	
Task performance	None	3.73(0.45)	**0.001**	3.03(0.72)	0.513	3.37(0.62)	0.197	3.13(1.07)	0.073	13.27(1.55)	**0.004**
	Any	3.17(0.73)		2.89(0.81)		3.13(0.77)		2.66(1.19)		11.85(2.26)	
Errors	None	3.17(0.79)	0.159	3.13(0.82)	0.723	3.37(0.81)	0.178	3.2(1.03)	0.135	12.87(2.1)	0.057
	Any	2.94(0.76)		3.09(0.8)		3.15(0.78)		2.83(1.19)		12(1.9)	
End-Product	None	3.47(0.63)	0.26	3.27(0.87)	**0.033**	3.63(0.56)	0.277	3.17(1.02)	0.151	13.53(2.01)	**0.021**
	Any	3.28(0.71)		2.81(0.95)		3.43(0.74)		2.77(1.22)		12.28(2.53)	
Column Total	None	13.97(1.65)	**0.008**	12.73(2.1)	0.167	13.87(1.43)	0.169	12.73(3.31)	0.149	**53.3(5.94)**	**0.015**
	Any	12.66(2.24)		12.04(2.27)		12.91(2.5)		11.74(3.6)		**49.36(7.6)**	

Table 4.12: L-TMEpt scores displayed for patients with and without any 30-day morbidity. Morbidity was associated with a significant four point decrease in mean score and a number of individual items, row and column scores. This result suggests surgical performance is associated with post-operative morbidity.

There was a significant difference of four points in the total score between cases with no morbidity and those were one or more events occurred within 30 days (53.3 (±5.94) vs. 49.4 (±7.6), p=0.015). Additional small but significant differences were observed in 3 of the 12 individual items, one column and two skill domains. Outside of anterior TME, the majority of p values were below 0.2 presenting the possibility of type II errors although the absolute differences are small.

When the three point case complexity grade data was considered, a clinically but not statistically significance increase in 30 day complication rates were present (48.6% vs. 67.9% vs. 78.6%, p=0.101), with more complications (median 0 vs. 1 vs. 2, p=0.058) and higher Clavien-Dindo grades seen in harder cases (median 0 vs. 2 vs. 2, p=0.027). There was no association between case

complexity score and anastomotic leak or re-operation (p=0.341 and p=0.872 respectively). Clinically but not statistically significant differences in length of stay was seen between the three grades (median 6 (IQR 5-13) vs. 8 (6-16) vs. 17 (7-14) days, p=0.095) and readmission rates (11.4% vs. 28.6% vs.28.6% p=0.186).

Attempts to define L-TMEpt score thresholds was hampered by the observed distribution of scores in patients that did and did not develop 30-day complications. Receiver operating characteristic curve did not support clinical utility of L-TMEpt scores (AUROC 0.335 (95%CI 0.213-0.458), p=0.015). OCHRA total and pelvic errors had higher areas on under the curve but insufficient for clinical practice (total case errors AUROC 0.713 (95%CI 0.589-0.837), p=0.002 and pelvic errors AUROC 0.66 (95%CI 0.532-0.751), p=0.787).

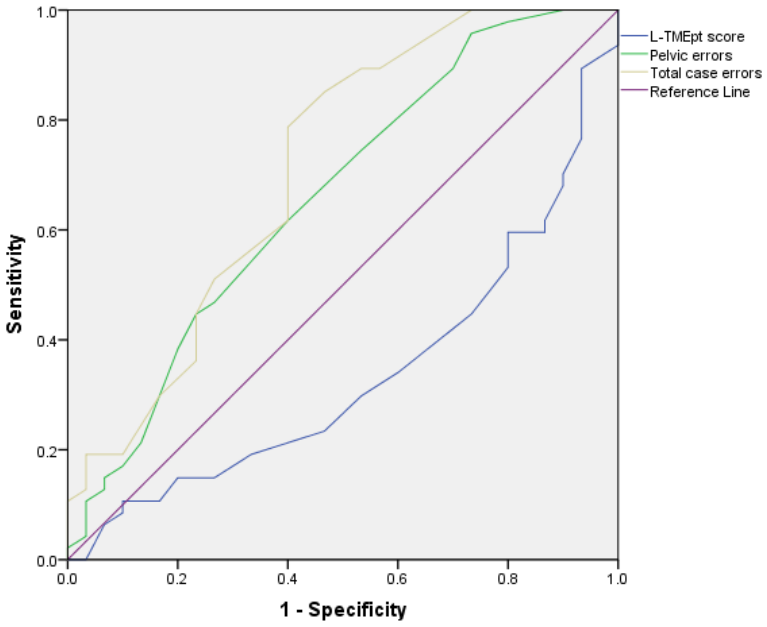


Figure 4.18: AUROC for 30-day morbidity. None of the objective assessment methods show utility.

4.11 Prospective application of intraoperative quality assurance to 2D vs. 3D surgery

When the objective assessment results were broken down into the 2D and 3D trial arms and compared, no differences were observed. Total cases errors were equivalent (18 (IQR 14–21) vs. 17 (IQR 13–22), p=0.437). For clarity, OCHRA categorical data tables and comparisons are not presented

here but provided in the publication (188). Across all comparison fewer overshoot errors in 3D cases was the only significant difference (64 vs. 48, p=0.05). All others were equivalent.

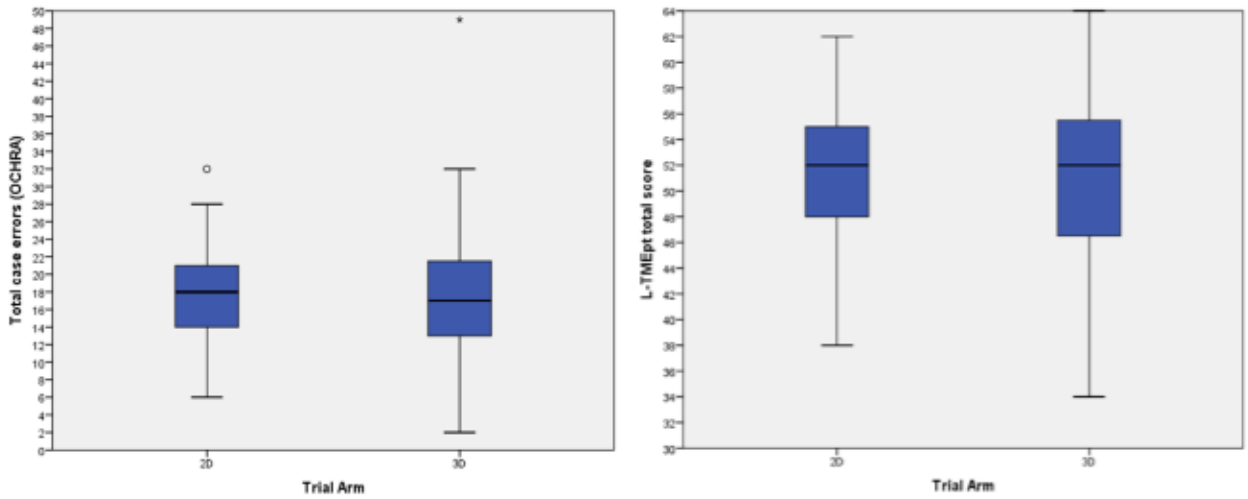


Figure 4.19: Box whisker plots comparing objective assessment results for 2D and 3D surgery. No differences are seen.

Due to its present research status, L-TMEpt data was not included in the published 2D3D manuscript (188). Tool scores were identical between 2D and 3D cases (52 (IQR 48-55, range 38-62) vs 52 (47-56, 34-64), p=0.886) with no difference in any individual item, row or column.

N=77	Trial Arm	Posterior TME		Anterior TME		Lateral TME		Resection and anastomosis		Row total	
		Mean (sd)	p	Mean (sd)	p	Mean (sd)	p	Mean (sd)	p	Mean (sd)	p
Retraction + Exposure	2D	3.32 (0.71)	0.319	3.27 (0.61)	0.936	3.35 (0.82)	0.568	3.38 (0.89)	0.982	13.32 (2.1)	0.877
	3D	3.48 (0.68)		3.28 (0.64)		3.3 (0.72)		3.4 (0.84)		13.45 (2.01)	
Task performance	2D	3.38 (0.68)	0.830	2.86 (0.67)	0.347	3.22 (0.71)	0.934	2.89 (1.05)	0.991	12.35 (1.86)	0.738
	3D	3.4 (0.71)		3.03 (0.86)		3.23 (0.73)		2.8 (1.26)		12.45 (2.35)	
Errors	2D	3.05 (0.74)	0.831	3.19 (0.84)	0.26	3.3 (0.7)	0.664	2.97 (0.93)	0.399	12.51 (1.68)	0.52
	3D	3 (0.82)		3.03 (0.77)		3.18 (0.87)		2.98 (1.31)		12.18 (2.32)	
End-Product	2D	3.38 (0.64)	0.849	2.86 (1.03)	0.359	3.57 (0.65)	0.469	3.08 (1.09)	0.242	12.89 (2.23)	0.869
	3D	3.33 (0.73)		3.1 (0.84)		3.45 (0.71)		2.78 (1.21)		12.65 (2.58)	
Column Total	2D	13.14 (2.04)	0.841	12.19 (2.2)	0.578	13.43 (2.14)	0.606	12.32 (3.14)	0.996	51.08 (6.47)	0.866
	3D	13.2 (2.21)		12.43 (2.25)		13.15 (2.25)		11.95 (3.84)		50.73 (7.93)	

Table 4.13: L-TMEpt score data for the 2D and 3D trial arms. No differences are seen.

When both objective assessment methods and trial arm data overlaid, beyond the weak negative correlation previously described, no clear pattern is displayed. The application of intraoperative quality assurance reveals that substantial measurable variation in performance is present across the whole trial. Despite the measures undertaken to try and provide two comparable groups, it may be considered that performance heterogeneity persists.

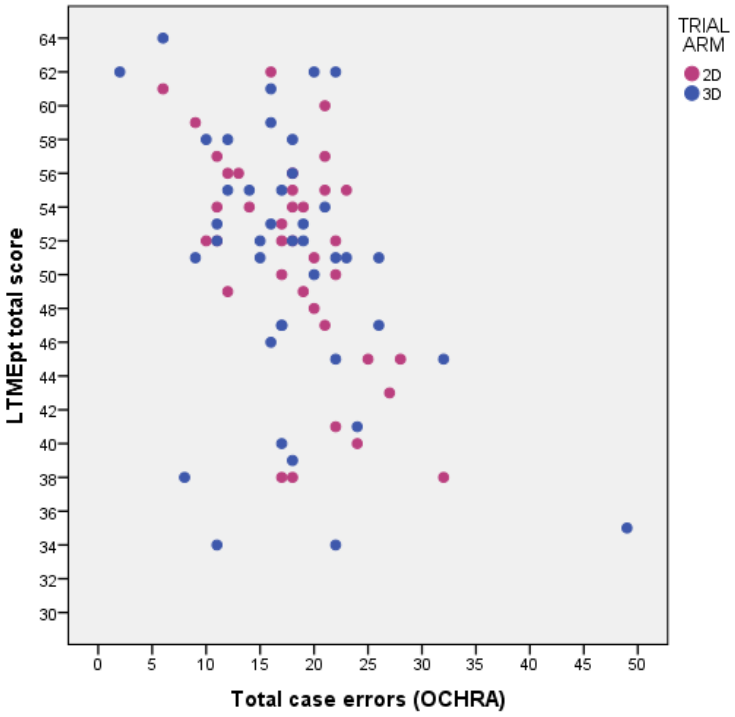


Figure 4.20: Scatterplot for objective assessment data for the trial arms. No pattern or clustering is evident with high and lower scores present in both groups.

4.12 2D vs. 3D laparoscopic TME RCT clinical results

This section reports the clinical and secondary outcomes of the trial.

Operative data

Median operative time was 270 minutes (IQR 240-348, range 120-640) with no differences between the arms (278 (270–360) vs. 270 (235–335), p=0.34). Similar results were seen for blood loss (median

80ml (IQR 50-150, range 0-1200) with no difference between 2D and 3D surgery (60 vs. 90 ml, p=0.618). There were four conversions to open surgery: three due to narrow pelvises preventing TME dissection and one due to difficulty identifying the tumour. Once more the arms were comparable (2 (4.9%) vs. 2 (4.8%), p = 0.981). Defunctioning ileostomy creation was equivalent (89% vs. 85%, p=0.587) as was anastomotic height (3 vs. 3 cm, p=0.829).

Case complexity – Surgeon reported and cognitive load

Using 100mm visual analogue scales with zero representing the easiest possible laparoscopic TME procedure, operating surgeons reported overall low case difficulty (median 29mm, IQR 18-47.5, range 1-93). Similar relatively low scores were observed across all other case steps with no difference in any comparison between 2D and 3D surgery.

Table 4.14 (next page): Surgeon reported difficulties. All figures represent millimetres. Surgeons marked 100mm visual analogue scales that ranged from 0 (easiest) to 100mm (hardest). For clarity no distribution data is presented. Overall low scores were returned with no difference in any comparison. 3D systems did not alter surgeon reported case difficulty.

All figures represent millimetres	2D Median	3D Median	p
Overall case complexity	28	31	0.399
Access to abdomen	14	13	0.784
Splenic Flexure mobilisation	21	18	0.127
IMA pedicle dissection and division	22	20	0.871
Access to pelvis	16	18	0.511
Identification of autonomic nerves	24	22	0.54
Division of rectum	19	20	0.919
Anastomosis	22	17	0.181
Anterior TME			
Anterior TME difficulty	30	25	0.78
Oedema	5	6	0.483
Fibrosis	8	8	0.327
Bleeding	6	8	0.4
Surgical Planes	14	13	0.838
Left lateral TME			
Left TME difficulty	19	22	0.705
Oedema	7	9	0.676
Fibrosis	7	10	0.363
Bleeding	9	10	0.86
Surgical Planes	14	16	0.68
Right lateral TME			
Right TME difficulty	25	30	0.29
Oedema	7	7	0.616
Fibrosis	10	14	0.316
Bleeding	9	12	0.504
Surgical Planes	20	20	0.38
Posterior TME			
Posterior TME difficulty	20	18	0.603
Oedema	7	6	0.524
Fibrosis	8	7	0.593
Bleeding	8	8	0.941
Surgical Planes	16	13	0.383

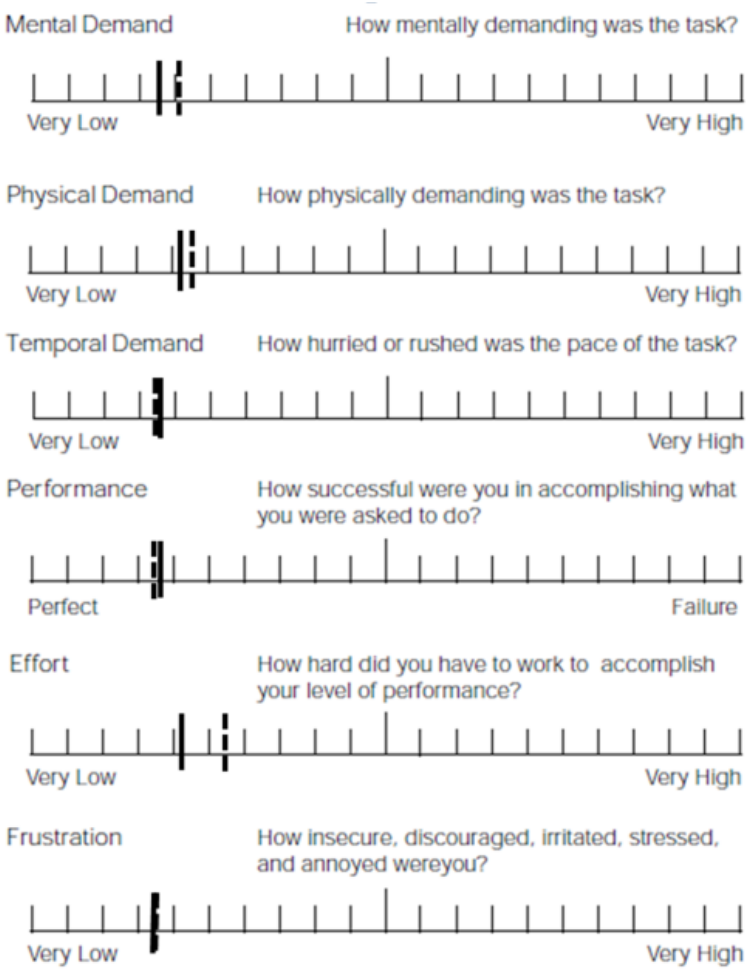


Figure 4.21: NASA task load index displaying surgeon reported cognitive load data. Medians denoted with dashed line (2D) and solid line (3D). For clarity interquartile ranges are not displayed.

Reported cognitive load was low across all six domains of the validated NASA task load index (203, 208, 209). No differences were seen in any component between the trial arms (p=0.186-0.952).

30 day morbidity

In the 85 patients that underwent surgery 110 morbidity events were recorded within 30 days of surgery (median 1, mode 0, IQR 0-2, range 0-5). 51 (60%) patients experienced any complication (median 2 events, IQR 1-3).

		Frequency	Percent
Number of 30 day morbidity events (per patient)	None	34	40
	1	17	20
	2	18	21.2
	3	10	11.8
	4	4	4.7
	5	2	2.4
	Total	85	100.0

Table 4.15: Number of 30 day morbidity events recorded in the trial. Figures represent individual patients.

The nature and degree of severity as defined by the Clavien-Dindo classification is displayed in table 4.9 (181, 183). 86.4% were minor (grade I and II). Seven patients returned to theatre (7.4%) including all five anastomotic leaks (trial leak rate 5.9%). 13.6% of morbidity events were grade III (surgical, radiological or endoscopic intervention required). Three patients (2.7%) required unplanned intensive care admission in the early post-operative period. There were no deaths.

Table 4.16 (next page): All morbidity events displayed by Clavien-Dindo grade and grouped by trial arm. Asterisk denotes a return to theatre occurred as a result of this complication.

Trial Arm Number of cases	2D 42				3D 43			
Clavien-Dindo Classification	I	II	III	IV	I	II	III	IV
Ileus	5	4			5	3		
Acute kidney injury	2	3			4	2		
Urinary retention	3				4	1		
Wound infection		5				1	1	
Sepsis		4				3		
Abdominal or pelvic collection		2	2			2	1	
High output stoma	1	1			1	3		
Urinary tract infection		4			1	1		
Atrial fibrillation, flutter or supraventricular tachycardia		3			1			1
Anastomotic leak			2**				3***	
Anaemia						2		
Hypertension					1	1		
Nausea/vomiting	1		1					
Stoma prolapse					2			
Pneumonia		1				1		
Splenic haematoma	1				1			
Allergic reaction					1			
Chest pain					1			
Diabetic ketoacidosis		1						
Duodenal ulcer bleed							1	
High output drain	1							
Hypocalcaemia						1		
Hypotension								1
Ischaemic optic neuropathy	1							
Neuropraxia					1			
Neutropenia							1	
Pancreatitis				1				
Rectal bleeding					1			
Retrograde ejaculation	1							
Small bowel obstruction			1*					
Stomal bleeding	1							
Stomatitis		2						
Vasovagal collapse	1							
Wound bleeding	1							
Sum	19	30	6	1	24	21	7	2
Total				56				54

		2D		3D		p
		Count	Column N %	Count	Column N %	
Tumour Stage	PathCR	0	0.0%	2 (22% Path CR)	4.7%	0.658
	1	15	35.7%	13	30.2%	
	2	13	31.0%	15	34.9%	
	3	13	31.0%	12	27.9%	
	4	1	2.3%	1	2.3%	
pT	PathCR	0	0.0%	2	4.7%	0.497
	1	4	9.5%	6	14.0%	
	2	18	42.9%	9	20.9%	
	3	18	42.9%	22	51.2%	
	4	2	4.8%	4	9.3%	
pN	0	28	66.7%	31	72.1%	0.687
	1	9	21.4%	6	14.0%	
	2	5	11.9%	6	14.0%	
pM	0	41	97.6%	42	97.7%	1
	1	1	2.4%	1	2.3%	
Relationship to peritoneal reflection	Above	22	52.4%	18	41.9%	0.188
	Astride	8	19.0%	6	14.0%	
	Below	12	28.6%	19	44.2%	
CRM (mm, median, IQR)		17.0 (10-25)		11.0 (6-18)		0.088
lymph node yield (median, IQR)		19 (15-27)		19 (14-26)		0.912
Plane of mesorectal excision	Mesorectal	32	76.2%	35	81.4%	0.163
	Intramesorectal	4	9.5%	1	2.3%	
	Muscularis Propria	4	9.5%	1	2.3%	
	Not reported	2	4.8%	6	14%	
R status	0	41	97.6%	42	97.7%	0.987
	1	1 (CRM 0.8mm)	2.4%	1 (distal margin <1mm)	2.3%	

Table 4.17: Pathological data for each trial arm. No differences are seen. PathCR=Pathological complete response to neoadjuvant chemotherapy.

Median length of stay was 7 days (IQR 6-16, range 2-96). As would be expected, patients with any complication stayed twice as long as those with any post-operative morbidity (median 6 vs. 12, $p<0.001$). Increasing numbers of 30 day morbidity events significantly extended length of stay ($p<0.001$). 16 (18.8%) were readmitted to hospital within 30 days of discharge with only one of the 34 patients without any morbidity event readmitted. The readmission rate in the 51 patients with any early morbidity was 29.4% ($p<0.001$).

4.13 Informing the design of future clinical 2D vs. 3D RCT

Although the aim of this chapter was to investigate the feasibility of applying intraoperative quality assurance within an RCT, the data obtained can also assist the design of future clinical trials on 2D vs. 3D laparoscopic TME surgery.

The delivery of this developmental RCT provides data to support the design of a future definitive RCT. Across all pre-defined endpoints, equivalence between the 2D and 3D trials arms was seen. The equal operative, cognitive load and patient outcome data suggests specialist performance was not altered by the imaging technology used.

The main finding of interest was the potential improvement in TME specimen quality following 3D laparoscopy. 95% of 3D TME specimens were assessed as mesorectal fascial plane representing a clinically, but borderline statistically significant difference to 2D surgery (80% vs. 95%, $p=0.059$). This figure exceeds the results reported by major laparoscopic rectal cancer trials as well as their open and robotic arms including where the primary endpoints was based on histopathology data (81, 85, 88, 89). This result and the low observed CRM involvement rate could be expected to lead to low rates of recurrence (39, 101). Therefore, TME plane of surgery is strongly advocated as the appropriate endpoint for definitive 3D laparoscopic TME trials.

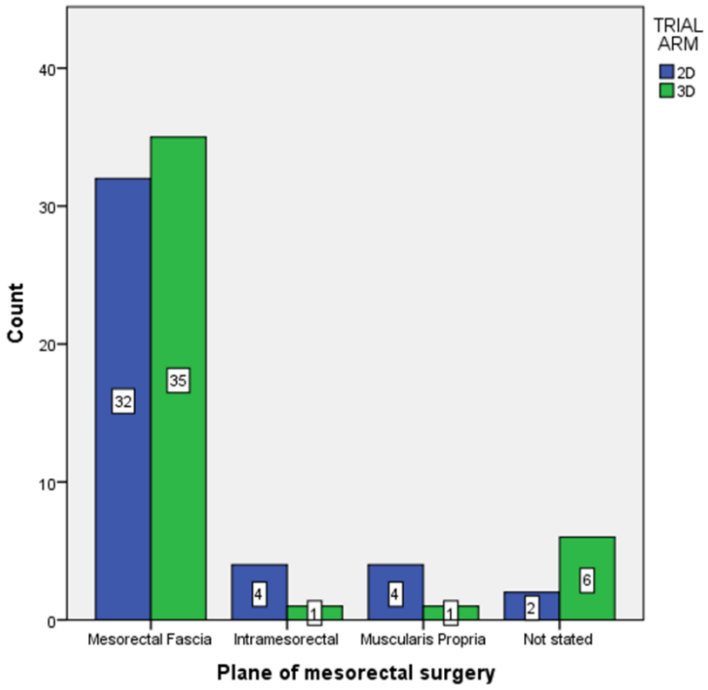


Figure 4.22: Bar chart for plane of TME dissection in both trial arms. Mesorectal fascial plane (80% vs.95%, p=0.059).

To maximise generalisability and not interfere with the clinical need for prompt tumour data to inform adjuvant treatment decisions, specimen assessments were made by local histopathologists. No plane assessment was made in eight cases and upon review no apparent reason is identified. The median tumour height was 11cm in these cases and it is possible that some were considered proximal to the mesorectum in keeping with common anatomical variability seen in the population (1, 4, 145) (section 1.1). This missing data may have altered the results and subsequent conclusions, but early identification of this issue shows the strength of undertaking preliminary studies which will directly improve future RCT design and likelihood of successful delivery. Retrospective specimen re-review was not possible following fixation, cut up and paraffin embedding for microscopic analysis. In hindsight, a photographic and/or pathology protocol with pre-trial training of participating pathologists and central data review may have reduced this issue and should be considered for future studies with requiring histology specimen reporting.

In a definitive trial, central histopathological assessment by credentialed expert assessors using a defined protocol represent stronger methodology through reduction a potential source of variation. This approach would require additional resources for specialist logistics to safely deliver specimens as well as ensure prompt reporting to meet clinical requirements. Central pathology review has been used successfully by a number of colorectal studies without delaying analysis but establishing a secure and reliable pathway would be mandatory (213, 214).

Power calculation for a definitive 3D laparoscopic TME trial

Using the plane of mesorectal surgery data, a number of sample size calculations for a future definitive trial were prepared with the assistance of the NIHR research design service.

Parameter	1	2	3	4	5	6
α	0.05	0.05	0.05	0.05	0.05	0.05
β	0.9	0.9	0.9	0.9	0.9	0.9
Proportion of 2D mesorectal fascia plane of surgery (X_1)	0.76	0.76	0.76	0.8	0.8	0.76
Proportion of 2D mesorectal fascia plane of surgery (X_2)	0.81	0.81	0.81	0.95	0.95	0.86
Allocation ratio	1:1	1:1	2:1	1:1	1:1	1:1
One or two sided	One	Two	One	One	Two	One
Group size 1	1155	1417	863	82	101	262
Group size 2	1155	1417	1726	82	101	262
Total	2310	2434	2589	164	202	524
Add attrition rate (12.5%)	289	304	324	21	25	66
Total sample size	2599	2838	2913	185	227	590

Table 4.18: Different sample size calculation options for a future definitive 2D vs. 3D TME RCT.

For all options, a power of 0.9 was considered appropriate for a definitive trial with α=0.05. Options 1-3 were based on intention to treat mesorectal plane data. The resultant sample sizes are not

deliverable. Alteration the allocation ratio to 2:1 with the aim of generating more 3D procedural data further expanded the figure and is therefore inappropriate.

Options 4-6 were based on the per protocol TME plane of surgery data. The larger difference corresponds to a greatly reduced sample size. As there is no concern that contemporary 3D platforms are worse than 2D laparoscopy, a one sided design is appropriate (193, 197). Options 4+5 would power the primary endpoint, but many practicing surgeons would not consider it as a definitive study. It is unlikely any useful secondary outcome data would be obtained and risks not furthering the knowledge gained from our developmental trial. Option six appears to fit requirements for a definitive 2D3D laparoscopic TME RCT. The number of patients would also provide important secondary outcome data including short term morbidity and health economics and long term locoregional recurrence, disease free, cancer specific and overall survival. If our primary outcome findings were reproduced, supportive data from these areas would be assist the drive to change practice.

We excluded abdominal-perineal and trans-anal TME excisions. Although the need for a complete specimen is unaltered in these procedures, this provided homogeneity for tool application. Variation in perineal and low rectal surgical technique, particularly where extra-levator surgery was performed could have directly influenced histopathology results (174). The subsequent sample size calculation would likely require the inclusion of APER, Hartmann's procedures and possibly trans-anal TME to reach the 590 target within a timeframe acceptable to a major research funder. In these operations the need for a high quality specimen remains unchanged as would the histopathological assessment undertaken justifying their inclusion. However, the quality assurance work presented in this thesis cannot be automatically assumed to apply to cases other than laparoscopic TME with anastomosis.

4.14 Discussion

Through the design, delivery and analysis of a bespoke multi-centre RCT of laparoscopic TME, intraoperative QA was successfully applied as the primary endpoint and used to assess all study

procedures. In this first MAS RCT to use this approach, the feasibility of embedding intraoperative QA is established and allowed the technical performance of credentialed consultant surgeons to be appraised. OCHRA and L-TMEpt results showed moderate inverse correlation with similar strength to the LapCo experience with colonic resections (73).

Within the context of the feasibility approach adopted and the associated study size this developmental study was not powered to answer the question of the superiority of stereoscopic imaging on the performance of Lap TME. As expected, no differences were seen between 2D and 3D surgery between any of the multiple endpoints. It has however highlighted a number of positive findings. Firstly, the feasibility of applying intraoperative QA into RCT across multiple sites is demonstrated with collection of meaningful data that could be reproduced by other trials including different disciplines. Secondly the study and measurement of intraoperative performance is shown to be possible and acceptable to patients and surgeons. The demonstrated heterogeneity in assessment scores is present despite all attempts to credentialing surgeons and standardise the study. It could therefore be considered that these steps are presently inadequate and could be augmented by intraoperative QA. Thirdly, the assessment data is seen to be clinically relevant and associated with important outcomes. Overall, the additional data provided by intraoperative QA argues for its wider application in RCTs. The usability and applicability features of the L-TMEpt are sufficient to support ongoing QA application with the notable advantage of alleviating the need for time intensive full video review.

The performance assessment data raises important questions. Together with the preceding chapter, substantial variation in the objectively assessed surgical performance is identified in both RCTs. Not only is it identified, the methods applied allowed it to be measured and quantified on a case by case basis. Within this small RCT such variation could have larger influence on results than larger trials.

In keeping with all RCTs, the 2D3D and ALaCaRT studies incorporated a number of design steps and accepted methodological frameworks to minimise bias aiming to establish homogenous trial arms

(144). This forms a key assumption as the differences in subsequent comparisons are believed to result from the intervention being studied. The performance variation reported here exists despite the attempts made by both trials including credentialing, quality assurance, video capture and study design considerations (89, 90, 188). The observed heterogenous performance levels may represent a highly important variable that has not previously been considered. Heterogeneity statistics are commonly applied and well understood within meta-analysis but equivalent assessments outside this setting are rarely considered.

The adopted objective assessment methods appear well suited to identify varying levels of performance. If construct validity were established within specialist surgeons previously believed to hold comparative skills sets this would hold a number of implications. Performance could account for outcome differences within surgical intervention trials and potentially those resulting from routine clinical practice. Objective performance data could further assist with interpretation of obtained results. Considering surgeon variability as a confounding factor to be accounted for in future trial design has methodological implications for future high quality randomised trials and their outcome data. Urgent assessment of this variation and its associated implications on both trial endpoints and clinically important outcomes is now indicated. This could form a new intraoperative quality assurance method and application of the methods adopted in this thesis. This work is undertaken in chapter five.

When short term morbidity is considered, a significant difference of four points in total L-TMEpt scores despite the homogenous level of the participating surgeons and patient and tumour demographics. Differences are also seen in a number of individual items and also between patients with no morbidity and those that developed any complication. Scores also significantly correlated with both the number and severity of these complications supporting both clinical and construct validity for this key patient centred outcome. This suggests that many post-operative events

originate from theatre performance further highlighting the need for high quality delivery of surgical interventions. Quality assurance could underpin this strategy.

The identified variation meant it was not possible to set either an OCHRA or L-TMEpt criterion referenced threshold. A broad range of scores were obtained for all TME planes and number of complications. Subsequently the sensitivity and specificities and areas under the curve confirm the L-TMEpt is insufficient in this form for clinical application particularly for specialist assessment.

Whilst acquiring case video for objective analysis, an additional research question was studied. There has been an uptake in 3D laparoscopy in clinical settings despite little prior evidence to support its use (193). The frequently overlooked intraoperative period was an appropriate point to study as it this is where any impact of imaging technology is most likely to be seen. It was hoped this approach could provide new insights into trial findings and identify areas for targeted improvement.

From the clinical perspective, as there was no prior 3D laparoscopy TME research, and as advocated by the IDEAL collaboration on surgical innovation, it was important to perform a developmental study in order assist the design of future definitive trials of this technology (139). Feasibility and multi-centre recruitment rates were also needed given the time and resource implications of major trials particularly those that collect medium to long term oncological follow up data. Incorporating all methodological recommendations for multicentre laparoscopic colorectal RCTs and 3D studies (110, 191, 192), this trial represents the first to investigate 3D laparoscopy within TME surgery.

As discussed in section 4.13, the main finding of potentially improved TME specimen quality with 3D surgery is of interest and may be a direct route towards quality improvement. The 95% mesorectal fascia result exceeds those seen in other TME RCTs and warrants ongoing exploration. This result also supports the use of MAS by the participating surgeons. As would be expected from the sample size no other differences were observed across the operative, surgeon reported, cognitive load, objective assessment of performance, early morbidity and pathological secondary outcomes studied.

Reflecting our exclusion of abdominal-perineal resections the average tumour height was slightly higher than the major trials and lower neo-adjuvant use was seen in keeping with UK guidelines (9).

The results of this developmental study are likely to be generalizable to other specialist laparoscopic rectal surgeons. Across all other pre-defined endpoints, equivalence between the 2D and 3D trials arms was seen. The equal operative, cognitive load and patient outcome data suggests specialist performance was not altered by the imaging technology used. It is possible their experience has overcome the lack of depth perception inherit to 2D laparoscopy. No meaningful surgeon side effects were encountered and no deterioration in cognitive load was seen suggesting contemporary 3D platforms have overcome past deficiencies (51). The trial was strengthened by the use of centralised randomisation with allocation concealment, standardisation of equipment across all the trial arms and all participant centres, stereopsis testing, blinded direct observation of performance and independent histopathology reporting and morbidity data collection.

Our study design was agreeable to patients, surgeons and theatre teams resulting in acceptable recruitment with low attrition which should be reproducible across additional sites. Should a definitive study confirm this finding it would represent an easily implemented and generalisable route to quality improvement whilst delivering the short-term recovery benefits presented by MAS (40, 86, 92). Outside this endpoint the equivalence of all other data does not support undertaking larger trials. Whilst the sample size presents the likelihood of type II errors the observed effects sizes appear small and of questionable clinical value to patients, surgeons and healthcare providers. The health economics of 3D laparoscopy have not been sufficiently reported to date although a recent health technology assessment suggested the cost per patient for 3D systems in general surgery could be as low as €1.67 (193). Our results suggest no meaningful secondary impact on healthcare resources could be expected.

The majority of 3D studies to date have used box trainer and laparoscopically naïve participants limiting the applicability to real world theatre performance and specialist surgeons (191, 192). The few clinical 3D randomised studies have also shown equivalent results including those performed some twenty years ago (215). It remains surprising that surgical technology undergoes intensive development and testing as part of the extensive requirement for licencing but clinical research assessment and prove of efficacy is not mandatory. This is in direct contradiction to the extensive regulatory requirements for other healthcare interventions particularly pharmaceuticals. Surgical intervention research presents challenges including those not frequently encountered in other clinical research areas but the need for evidence-based practice remains including the selection of theatre technologies (139, 216, 217). It is noteworthy that the trial surgeons repeatedly praised stereoscopic vision and subjectively felt their performance and outcomes were superior with 3D. Many were highly surprised when data was unblinded and results revealed. This effect has been noted by other colorectal MAS technology trials where robustly performed studies failed to show the results hoped by advocates (81, 218). The CONSORT criteria can be considered as a QA framework for RCTs. Randomisation removes many of the inherent biases that can unduly influence comparable studies and is considered the default option for surgical intervention and technology studies by the IDEAL collaborative (139, 216, 219). The results obtained reflect the expertise of the participants and cannot be assumed to be applicable to trainees or inexperienced laparoscopic surgeons. It may be that 3D laparoscopy only benefits the trainee level.

4.15 Critique and study limitations

The original purpose of the L-TMEpt may be responsible for the observed results. Based on the clinical need reported by the literature base, like the ALaCaRT trial the 2D3D study was composed of specialist surgeons. The L-CAT was designed for the training setting and the L-TMEpt utilised similar methodology in its development. It is noteworthy that in the original 12 patient L-TMEpt pilot, operated by four UK consultant colorectal surgeons and assessed by six experts, almost identical

findings were seen (129). Further evidence is shown by a probable ceiling effect with consistently high tool scores and the vast majority of items scored as 3 or 4. Therefore the data appears somewhat clustered and may not fully differentiate between differencing levels of performance. This result however is reassuring suggesting high quality surgery is being delivered. If a wider distribution was seen it would present potential ethical concerns with high stakes implications for patients, hospitals and individual surgeons.

Video analysis provides a comprehensive review of laparoscopic interventions (185). However extra-corporeal tasks are not captured by this strategy meaning errors and their associated impacts and lessons may have been missed. Additionally, non-human factors are not captured such as distractions, team composition (particularly the experience of the assistant surgeon) and familiarity which could all potential influence performance (220). Although the study design maintains face validity and generalisability to UK practice, and we used the widely applied NASA-TLX to capture cognitive load and asked surgeons to assess case difficulty there is a risk that external factors with possibility implications on performance were missed. Direct error comparison with other procedural datasets is prevented by the widely differing methodologies and error definitions used in the literature (221). However measured, errors appear commonplace in laparoscopic surgery (222).

Across all centres, completion of case screening logs was poor and so it was not possible to report data regarding total potentially eligible patients, numbers approached and TME procedures performed open or robotically during the study timeframe. Whilst this would not have altered our recruitment rate data its absence goes against the CONSORT principles and reduces transparency. Screening logs would not be expected to capture pre-operative decision making processes and other factors such as the availability of a laparoscopic surgeons that may influence the surgical approach offered for each case. Patient fitness and wishes, local excisions, enrolment into other CRN portfolio trials, surgeon availability, cancer pathway timelines and provision of training to local registrars are some of the factors it was not possible to control for. Therefore, selection bias cannot be excluded.

Additionally, there was no difference in the number of OCHRA identified mesorectal injury errors and the histopathologists TME quality grade. This surprising finding may be accounted for by the definitions used in the Royal College of Pathologists guidelines (14). For example, 5mm deep injuries into mesorectal fat still permit an overall mesorectal fascia grade to be awarded whereas such an event(s) would be captured by OCHRA. It could be argued the TME grade definitions retain an element of subjectivity and to my knowledge no inter-rater reliability studies have been reported.

OCHRA case review is a time intensive undertaking which limited assessment to a single assessor. Although this same issue limits routine use, single assessment could have implications for high stakes assessments and those involving specialist surgeons with potential patient care and healthcare provider implications. However, single observer OCHRA assessment has previously been shown to be valid and reliable (223). The single assessor and the absence of a longitudinal aspect prevents the application of advanced statistical techniques such as generalisability theory, exploratory factor analysis and source of variance analyses. These can be applied to define the importance of various factors on obtained results and would be crucial before adoption of these methods is considered for high stakes assessment (224).

Given the research question and sample size, the 2D3D trial did not collect long term outcome data. Three and five year oncological data are required in future larger studies but unless the effect size of 3D surgery was seen to be large, it is unlikely that any significant long term differences would be observed, and the required resource implication may not be justified. Considering the three patient difference in mesorectal fascial surgery observed here, it is highly unlikely that this would equate to a meaningful difference in oncological outcomes.

Typically, samples size calculations and comparative statistic tests including p values are not appropriate when reporting feasibility studies (206). They are included here as they were considered requested by ethics review boards, study funders, trial sponsors and journal peer-reviewers. As the study was not powered to find any difference (other than a 30% difference in errors), any observed

differences could be considered as type I errors and it could be argued that identifying a statistically significant result negates the need for future research. Additionally, where differences are not shown, it cannot be assumed that this is true due to the underlying risk of type II errors. A number of p values (particularly table 4.12) were below 0.2 although the absolute differences are small.

Although we successfully met the aims of this trial, as a developmental study with a modest sample size, firm conclusions on the role of 3D laparoscopy should not be drawn. With the increasing clinical use of 3D systems, a health technology assessment report and a European consensus conference have recently reported (193, 197). Both support use of the technology with a small but significant reduction in operative time (8-15%) but the striking meta-analysis finding of a 50% reduction in complications is noted and clearly deserves dedicated further study. No difference was seen in the 2D3D RCT although overall morbidity levels were higher than the comparable literature. Data was prospectively collected by UK CRN portfolio funded specialist research nurses independent of the trial who visited each patient daily. As shown in table 4.8 detailed data was obtaining including many events that would not routinely be reported by surgical trials. Their inclusion is justified as a transparent and patient centred approach. Discounting these events as they are typically either considered unimportant or not directly linked to surgery appears simplistic as they impact upon the patient and their journey. A qualitative investigation of past patients could explore this area further.

4.16 Conclusions

Within a bespoke RCT the prospective application of objective intraoperative assessment of laparoscopic TME surgical performance was successful. Substantial variation in the technical skill of credentialed surgeons delivering cancer intervention within a multicentre RCT is evident and can be objectively measured. This variation may influence trial results and clinically relevant outcomes and requires further investigation. Combination of the two RCT datasets may permit this.

The feasibility of performing multicentre 3D laparoscopic multicentre trials of specialist performed complex surgical interventions is established. Provision of stereoscopic imaging was not seen to alter

laparoscopic TME performance or clinical and histopathological endpoints. There is a potential improvement in mesorectal specimen quality with 3D laparoscopy which should form the focus of future TME trials.

Chapter 5 – Can quality assurance measure performance variability and associated clinical impacts?

The application of intraoperative quality assurance reported in the preceding two chapters revealed that a detectable and measurable variation in surgical performance was present within the two RCTs. In this chapter combination of the two multicentre trial datasets is performed to provide a large dataset explore this further and investigate any implications arising from this observation.

Access to the ALaCaRT dataset was only possible with the kind support of Associate Professor Andrew Stevenson (Chief Investigator, Brisbane, Australia) and ALaCaRT trial management group. Written approval for this sub study was obtained from the Australasian gastrointestinal trials group (AGITG) who sponsored the trial. Combining the datasets was performed with the expert assistance of Mr Chris Brown (ALaCaRT Biostatistician) and Prof Val Gebski (Biostatistics and research methodology), National Health Medical Research Council Clinical Trials Unit, University of Sydney.

5.1 Introduction

High quality surgery underpins optimal patient and oncological outcomes following TME surgery as well as other surgical oncology procedures. Although this principle is accepted, considerable variation has been seen in results reported by randomized controlled trials (RCT) involving cancer surgery (40, 225). Surgical interventions are often grouped together for analysis as they are perceived as homogenous groups. Using right hemicolectomies as an example highlights the drawbacks associated with this strategy as surgeons may alter their distal resection margin based on variable factors such as tumour location, vascular anatomy and expected nodal drainage pattern. Complete mesocolic excision procedures necessitate more extensive central and longitudinal dissections providing further sources of heterogeneity. Therefore, grouping interventions risks oversimplification of inherent operator and procedural variability which could be magnified in more complex interventions (139).

TME represents an advanced, technically complex procedure requiring precise dissection within the narrow confines of the deep pelvis. As previously discussed, oncological outcomes following TME are intrinsically linked to the performance of surgery and the quality of the resection specimen (41, 76, 77, 101). As shown in section 1.6, within the laparoscopic TME RCTs complete resections were obtained in the majority of patients with reasons behind suboptimal outcomes remaining unclear (40, 93). This was also observed within the 2D3D and ALaCaRT analyses presented in this thesis.

It is a logical and widely held assumption that surgical skill influences procedural delivery and subsequent outcomes. In the few available reports, technical skill as was measured through blinded peer observation of laparoscopic gastric bypass performance using the validated objective structured assessment of technical skill (OSATS) system, was closely associated with a host of clinically important outcomes (130). This seminal study was the first to measure, quantify and compare objectively measured specialist performance within theatre environments and show that significant variation exists and holds direct intra and post-operative clinical importance.

There is a more compelling argument to investigate this within the surgical oncology setting, as the delivery of complex interventions may be subject to wider variation with potentially larger outcome effects. If present, this partial homogenisation between the trial arms will undermine the use of randomisation and differing interventions. The objective quality assurance analyses performed have identified variation within the two studied randomised trials. The clinical impacts, implications and acceptability of this identified specialist performance variation now requires evaluation. In order to achieve this a larger sample is indicated and is achievable by combining data from both RCTs.

Variation in surgical performance could account for the reported outcomes differences following laparoscopic TME surgery (40, 102, 225). In keeping with the overarching aims of this thesis, an exploration into the clinical implications of this observation is now indicated.

Aim

To assess whether the variation in intraoperative laparoscopic TME quality assurance results holds associated clinical implications.

5.2 Methods

Combining ALaCaRT and 2D3D datasets

Combination of the trial databases was greatly assisted by the use of identical variables and standardised definitions applied by both trials for the majority of fields. No patient identifiable data was shared between each trial group during this exercise.

It was not possible to combine the following areas due to non-compatible measurements and/or the inability to accurately apply one method to the other trials source data: patient fitness (ALaCaRT – Eastern Cooperative Oncology Group/World Health Organisation performance status; 2D3D American Society of Anaesthesiologists physical status score), anastomosis height (not recorded in ALaCaRT), relationship to peritoneal reflection (not recorded in ALaCaRT), whole case OCHRA data

(ALaCaRT videos only contained pelvic surgery). These factors were therefore dropped and not further analysed.

Whilst ALaCaRT recorded the number of complications as well as specifically recording pre-defined major complications including anastomotic leaks and re-operations, no additional data was held on events considered as minor complications. Severity of each complication was graded with the common terminology criteria for adverse event reporting system (version 3.0) (https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm). There is considerable overlap with the Clavien-Dindo classification applied within the 2D3D trial. The original 2D3D source data was reviewed and where required amendment was made following joint agreement between the Chief Investigator and myself.

Combined video analysis

As stated in section 3.6, following the ALaCaRT protocol the majority of cases did not capture sufficient resection and anastomosis tasks to allow full application of the L-TMEpt. To enable trial combination and homogenous video library, analysis of the first three TME dissection task columns was performed (TMEdt) as described in section 3.6. The possible TMEdt score range was 12-48.

As no differences were seen between the demographics, operative data, cognitive load, case difficulty, histopathology and early morbidity data between the 2D and 3D trial arms (section 4.12), the trial was merged for this analysis. When combined with the ALaCaRT videos, 176 cases with clinical data were available for analysis representing the largest objectively analysed surgical case video series of any procedure to date.

Clinical, construct and concurrent validity analyses

The approach of the Birkmeyer study was replicated to compare differing levels of performance by grouping surgeons into quartiles (130). Individual surgeons were labelled with their personal mean TMEdt score which was calculated as the sum of TMEdt scores from all included cases divided by number of analysed cases. These were ordered and used to divide them into lower, interquartile and

upper quartiles. No surgeon identifiable data was present in the database as both trials applied a unique surgeon ID as sole identifier maintaining anonymity.

Clinical endpoints comprised 30-day patient outcomes with morbidity events quantified and graded with the common terminology criteria for adverse event reporting system (v3.0). Re-operation, anastomotic leak, length of post-operative hospital stay, and readmission were specifically analysed. Subjective surgeon reported case difficulty was captured in the 2D3D trial using a 100mm visual analogue scale which was compared with L-TMEpt case complexity grades (0mm – easiest, 100mm hardest case) (188). Histopathological outcomes were plane of mesorectal surgery as reported by blinded pathologists (76, 77), the ALaCaRT composite primary endpoint (mesorectal fascial plane, circumferential margin ≥1mm and distal margin ≥1mm) and lymph node yield. Medium term oncological outcome data was captured for both trials including length of follow up, recurrence (any, locoregional, distant and time to relapse) and overall survival.

TMEdt scores were compared with pelvic errors identified with OCHRA in keeping with the description provided in section 3.5.

Statistical analysis and data handling

TMEdt scores were handled as a continuous variable with the assumption that all items carried equal weight. As mentioned above, to ensure homogeneity presented analyses represent TMEdt total with possible scores between 12 and 48. Clinical validity comparisons were performed using Mann-Whitney U, Kruskal-Wallis and Spearman rho correlation testing as appropriate. Number needed to treat (NNT) was calculated as the inverse of the absolute risk reduction (lower quartile outcome % - upper quartile outcome %). Unless otherwise stated data is displayed as median (interquartile range). Analyses were performed using SPSS™ (v25.0 IBM, USA) with $p<0.05$ considered significant.

5.3 Results

176 patients underwent laparoscopic TME with complete data capture including video recorded operations. Demographics and tumour characteristics for the combined cohort are displayed. No discernible or statistical difference was seen when compared to each individual trial (table 4.5) (89).

		Mean (SD)	Column %
Age		66 (12)	
Body mass index (kg/m^2)		27 (5)	
Sex	Female		40.0%
	Male		60.0%
Neoadjuvant Treatment	No		61.1%
	Yes		38.9%
Tumour Height from the anal verge (cm)		8.2 (3.3)	
Tumour Height	Upper (10.1-15cm)		24.0%
	Mid (6.1-10cm)		54.3%
	Lower (≤6cm)		21.7%
Tumour stage	Path CR		1.1%
	I		28.0%
	II		25.1%
	III		43.4%
	IV		2.3%

Table 5.1: Combined cohort demographics and tumour characteristics.

Clinical Validity

There was evidence of substantial TMEdt score variation (40 (36-43), range 25-48, figure 5.1). There was no difference in scores between the trials (40 (36-42) vs. 41 (36-44), p=0.187).

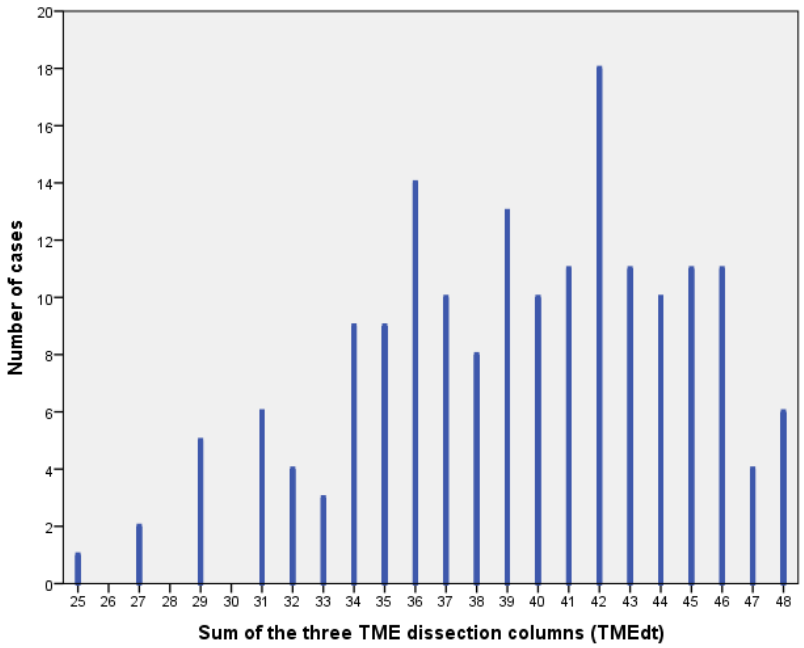


Figure 5.1: TMEdt score distribution for the combined trials dataset. A wide score range is evident.

Mean (SD)	Posterior TME Dissection	Anterior TME Dissection	Lateral TME Dissection	Resection and anastomosis	Total
N=	176	176	176	91	91
Retraction + Exposure	3.41 (0.69)	3.24 (0.69)	3.30 (0.7)	3.42 (0.82)	11.76 (2.64)
Task performance	3.38 (0.66)	3.07 (0.76)	3.29 (0.69)	2.92 (1.13)	11.26 (2.34)
Errors	3.11 (0.82)	3.29 (0.8)	3.39 (0.71)	3.08 (1.11)	11.38 (2.29)
End-Product	3.38 (0.69)	3.05 (0.94)	3.49 (0.67)	3.03 (1.15)	11.48 (2.61)
Column Total	13.28 (1.93)	12.65 (2.16)	13.49 (1.99)	12.45 (3.36)	TMEdt (n=176) 39.43 (5.07) L-TMEpt (n=91) 51.95 (7.24)

Table 5.2: L-TMEpt item, column, row and total with TMEdt total. Figures represent mean and standard deviations.

Case complexity was: grade 1 61 (34.7%), grade 2 85 (48.3%) and grade 3 30 (17%). Significantly lower TMEdt scores are seen with each increase (43 (40-46) vs. 39 (36-42) vs. 36 (32-38), $p<0.001$).

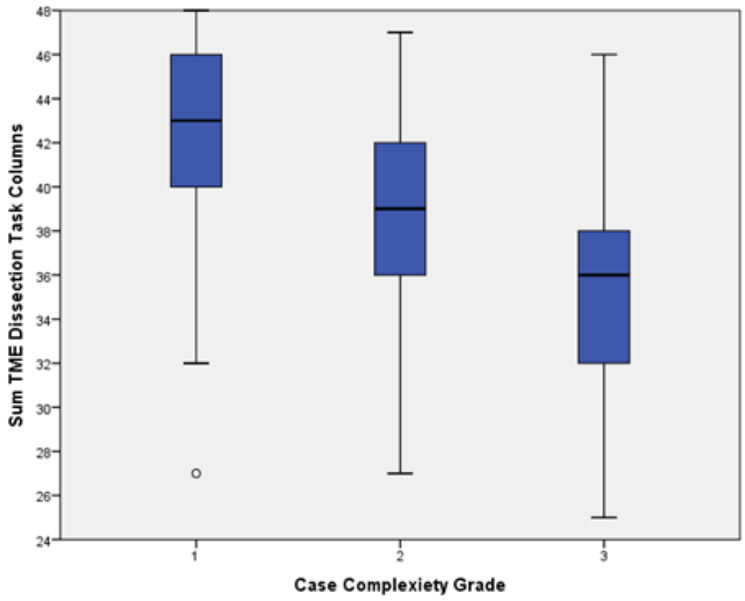


Figure 5.2: Box and whisker plot for TMEdt scores and grade complexities. Boxes represent medians and interquartile ranges with whiskers covering the 98% confidence interval. Scores significantly decrease with increase case grade.

Concurrent Validity

In 381 hours of unedited pelvic video analysis, 1115 pelvic errors were identified (median 5 per case (3-8), mode 3, range 0-31). There was moderate negative correlation with TMEdt (r^s=-0.515, p<0.001).

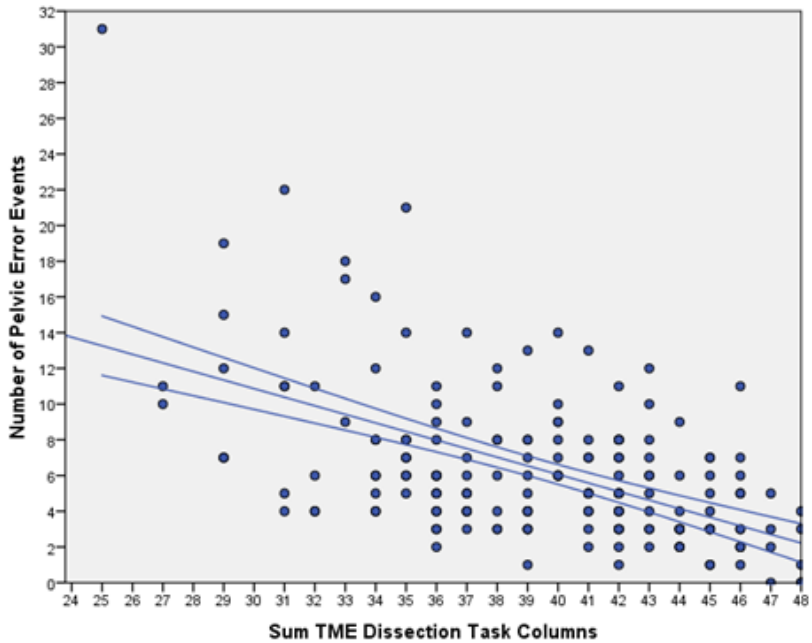


Figure 5.3: Scatterplot depicting OCHRA pelvic errors and TMEdt score data for the combined trial dataset with line of best fit and 95% confidence interval.

Construct Validity

Individual surgeon mean TMEdt scores varied between 29-48. The median was 39.3 and the interquartile range used to define the upper and lower quartiles was 42 and 37.3 respectively. Surgeon quartiles were applied with outcome data displayed (table 5.3).

Patient demographics were not equally distributed between the quartiles. Top quartile surgeons operated on patients with lower, more advanced rectal cancers that received more neoadjuvant treatment. Despite this, their cases were performed significantly faster with less blood loss and fewer pelvic error events (6 vs. 6 vs. 3, $p<0.001$).

Surgeons outside the lower quartile achieved significantly higher mesorectal fascial plane rates (93.3% vs. 88.9% vs 58.8%, $p=0.002$; NNT 2.9) and ALaCaRT composite endpoint success (88.3% vs. 88.9% vs. 58.1%, $p=0.032$; NNT 4). There was no difference in lymph node yields.

85 (48.3%) of patients developed a morbidity event within 30 days of surgery. Patients operated by top quartile surgeons developed significantly less 30-day morbidity (23.3% vs. 48.4%, $p<0.001$; NNT 4) as well as fewer and less serious events. A number of clinically relevant reductions in anastomotic leak (6.5% vs. 0; $p=0.161$; NNT 15.4), length of stay (9 (5-14) vs. 7 (5-11) days, $p=0.2$) and hospital readmission (33% vs. 0, $p=0.186$; NNT 3) were seen between the upper and lower quartiles but none were statistically significant.

		Surgeon Quartiles									Three groups p	Lower-upper p
		Lower			Interquartile			Upper				
		Median	IQR	%	Median	IQR	%	Median	IQR	%		
Performance measurement												
Laparoscopic TME performance tool score (TMEdt)		36	32-39		40	36-43		44	42-46		**<0.001**	**<0.001**
Pelvic OCHRA error count		6	5-11		6	4-8		3	2-4		**<0.001**	**<0.001**
Demographics												
Age (years)		68	61-77		67	60-75		6	55-70		0.134	0.066
Body Mass Index (kg/m2)		27	25-31		27	24-30		25	22-28		0.169	0.094
Tumour Height (cm)		8.0	7.0-10.0		8.0	6.0-11.0		7.0	5.0-8.6		0.119	**0.036**
Sex	Female			51.6			36			43.3	0.267	0.521
	Male			48.4			64			56.7		
Neoadjuvant Treatment				35.5			33.3			63.3	**0.01**	**0.031**
Tumour stage (Histopathologically defined)	PathCR			0			1.8			0	0.109	**0.041**
	I			38.7			25.4			26.7		
	II			29			28.1			10		
	III			32.3			42.1			60		
	IV			0			2.6			3.3		
Operative data												
Operative duration (minutes)		290	230-380		255	204-297		178	155-210		**<0.001**	**<0.001**
Estimated blood loss (ml)		100	50-200		100	50-200		40	15-60		**<0.001**	**<0.001**
Histopathological outcomes												
Plane of mesorectal dissection	Mesorectal Fascia			58.6			88.9			93.3	**<0.001**	**0.002**
	Intramesorectal			27.6			7.4			3.3		
	Muscularis Propria			13.3			3.7			3.3		
ALaCART Composite Endpoint Success				58.1			88.6			83.3	**<0.001**	**0.032**
Lymph node yield		16	11-27		17	13-22		17	12-21		0.905	0.665

Clinical Outcomes												
Any 30 day morbidity event				48.4			55.3			23.3	**0.008**	**0.043**
Number of 30 day morbidity events	None			51.6			44.7			76.7	**0.013**	**0.049**
	1			16.1			24.6			10		
	2			25.8			14			6.7		
	3			3.2			9.6			6.7		
	4			3.2			5.3			0		
	5			0			1.8			0		
Highest grade morbidity experienced (Common terminology for adverse event reporting system)	None			51.6			44.7			76.7	**0.024**	0.07
	1			6.5			10.5			3.3		
	2			29			28.9			10		
	3			12.9			11.4			6.7		
	4			0			4.4			3.3		
Unplanned Re-operation				3.2			1			3.2	0.599	0.981
Anastomotic Leak				6.5			5.3			0	0.406	0.161
Length of stay (days)		9	5-14		7	6-13		7	5-11		0.456	0.2
Hospital Readmission				33.3			18.2			0	0.228	0.186

Table 5.3: All data fields displayed for lower, inter and upper quartile groups with comparisons presented for upper vs. lower and three groups. Multiple major outcome differences with clinical implications are evident.

Oncological outcomes

Median follow up was 2.9 years. Locoregional recurrence was 4%. There were no differences in medium-term locoregional or distant recurrence or survival data between the quartiles (table 5.4). Although not statistically significant, clinically important findings were observed in patients operated by upper quartile surgeons with no locoregional recurrences, over two years longer disease-free survival and 96.6% overall survival.

	Surgeon Quartiles										
	Lower			Interquartile			Upper			Three groups p	Lower-upper p
	Median	IQR	%	Median	IQR	%	Median	IQR	%		
Follow up											
Follow up (years)	2.28	0.86-4.05		2.6	1.71-4.07		4.12	3.09-5.03		<0.001	0.003
Recurrence											
Disease Free Survival			74.2			78.1			70	0.634	0.717
Time from surgery to first recurrence (years)	1.67	1.39-4.49		2.92	1.58-4.14		3.98	1.44-4.53		0.667	0.514
Locoregional recurrence			6.3			4.5			0	0.435	0.175
Distant recurrence			21.9			13.4			24.1	0.269	0.835
Overall Survival											
Overall Survival			87.5			87.5			96.6	0.459	0.202

Table 5.4: Oncological outcome data displayed by surgeon quartile.

Case score quartiles (TMEdt)

An analysis based on individual case TMEdt score quartiles (irrespective of the operating surgeon and their mean TMEdt score) is presented in appendix 2. Similar results are observed although the effect size appears smaller for a number of outcomes (mesorectal fascial plane 76% vs 97.4%, NNT 4.7, p=0.006; ALaCaRT composite endpoint success 75.5% vs. 92.9%, NNT=5.7, p=0.025 and any 30 day morbidity 54.7% vs 38.1%, NNT=6, p=0.109, appendix 2).

Full L-TMEpt score data

Given the design of the tool, data from complete tool application is indicated. 87 laparoscopic TME case videos were seen to capture all resection and anastomosis tasks allowing full application of the L-TMEpt (14 ALaCaRT and 73 2D3D). For clarity, this analysis and table are presented in appendix 3.

5.4 Discussion

There is significant variation in outcomes following cancer surgery (40, 41, 77, 102, 225). The underlying reasons and their potential impacts are not clearly understood as surgical performance is not robustly measured. The aim of this chapter was to merge the trial datasets in order to assess the clinical implications arising from identified surgical skill variations within the RCTs to answer my second hypothesis. The relationship of surgical performance within complex cancer interventions and subsequent clinical and pathological outcomes has not been previously investigated.

Study of the combined data from the two trials confirmed, that despite the credentialing methods adopted, the identified variability in surgical skill persisted among specialist surgeons performing laparoscopic TME across four countries supporting the generalisability of this finding. Application of the L-TMEpt allowed this to be measured and quantified facilitating comparisons. Although this evaluation was performed on a large group of surgeons believed to hold equivalent skill levels within RCT settings our findings are likely to be applicable outside clinical research. The six minute average L-TMEpt completion time supports ongoing use and assists efforts to move away from labour intensive methods such as OCHRA which has comfortably exceeded 1800 analysis hours during this research.

The relationship between the identified performance variation and clinical outcomes was explored. Resulting surgeon and individual case score quartiles were strongly associated with perioperative data with upper quartiles delivering faster procedures that contained fewer errors despite a significantly higher proportions of demographic and tumour features traditionally considered to increase case difficulty. The benefits of high performance extended into improvements in the key histopathological determinates of oncological outcomes (41, 76, 77, 101). Large effect sizes were observed in rates of mesorectal fascial plane surgery that translate to exceedingly small numbers needed to treat. For just every three cases performed, upper quartile surgeons would obtain an extra mesorectal fascial plane specimen over their lower quartile colleagues. This research shows the first direct link between objectively measured surgical performance with important patient, clinical and cancer outcomes. I am

unaware of any other intervention in the multidisciplinary care of rectal cancer patients presenting such a large effect size with the potential for beneficial quality improvement.

Regarding short-term morbidity, similar patterns were seen with twice as many patients operated by lower quartile surgeons developing morbidity compared to the upper quartile group. They additionally developed more events with higher severity gradings. Upper quartile surgeons would need to treat just four patients to avoid morbidity that would be expected after surgery from a colleague in the lower quartile. Morbidity has the potential to be influenced by multiple factors including but not limited to patient co-morbidity and variations in peri-operative care. This study was based on performance from dozens of centres preventing any control of these factors. Despite this the intraoperative phase is shown to be a key determinate of post-operative outcomes.

In the assessment of surgical care, the intraoperative period receives comparatively little scrutiny with postoperative metrics typically reviewed. Our findings strongly suggest that operating theatre events should not be overlooked, and it could be further argued that surgical quality improvement initiatives should be refocused to concentrate on this period. The measurement of human performance particularly within complex systems can be challenging but the experience presented here shows it is possible and can add novel, clinically relevant and applicable information not obtainable elsewhere.

Variation

Although one might consider that it is to be expected that performance influences outcomes, this study is the first to demonstrate a direct link between surgical skill variation at the specialist level and a subsequent impact on outcomes. In the surgical oncology setting surgeons are shown to constitute an important outcome variable. Across the trials, measurable substantial inter-surgeon variation in the operative performance of complex laparoscopic interventions is demonstrated despite attempts to prevent this. It was not possible to define a score threshold for acceptable performance as sensitivities and specificities were lowered by a number of good clinical outcomes after low score cases and vice versa.

Moving away from surgical contexts, variation can be considered to exist in just about every aspect of life. Furthermore, these differences are to be expected at all performance levels. Using sport as an example, there are obvious differences between the elite and club levels as well as within each setting. Yet at all levels, coaching, training and practice are employed to try and improve with the belief that this is both possible and achievable. It is not clear if such methods can effectively drive improvement in specialist surgical performance of complex procedures or if some surgeons are simply better than others. However, both the need and opportunity for improvement is displayed in the results of this chapter and wider thesis and it is hoped this inspires surgeons and researchers in this field.

Comparison of results with the laparoscopic TME evidence base

In the context of the present debates on the oncological safety of minimal access techniques in rectal cancer, our data shows acceptable results, comparable and arguably superior to those reported following open TME, were obtained by surgeons above the lower quartile. Top performing surgeons were seen to have higher levels of patients and tumour factors typically considered to increase difficulty but still obtained excellent results. The laparoscopic vs. open TME RCTs all investigated the generalizability of laparoscopic techniques but in light of this thesis it is also appropriate to consider who is performing the procedure and their objectively demonstratable skill level rather than just the surgical approach undertaken. It would be fascinating if other reported laparoscopic vs. open TME trials could be revisited and re-analysed with the approach reported here although as outlined in section 3.2 this is prevented by the absence of operative videos. Considering the surgeon and their performance as a variable factor may have altered the conclusions reached and subsequent practice.

Implications for surgical intervention studies

These results have implications for methodologists and trialists performing studies on or containing complex surgical interventions. It is shown that surgical skill can be successfully measured, and that variation is present. Therefore, objective assessment of surgical performance seems appropriate for consideration by future clinical studies containing surgical interventions (3). Here, both trials credentialed their participating surgeons primarily based on prior case numbers, but this alone now

appears insufficient as heterogeneity persisted. Inclusion of objectively measured procedural competency could be implemented as an enrolment requirement. Presently subjective assessment of pre-trial videos by steering committee members is occasionally used for credentialing or quality assurance purposes (110). Tool application could provide objective and reliable performance evaluation and facilitate standardisation and comparison between assessors, surgeons and units. Individual case factors require consideration as correlation between scores and case complexity grades was observed. It is unclear whether a well-performed operation makes the procedure appear less complex, or if a less complex procedure allows a surgeon to demonstrate better performance. It was not possible to define a threshold for high stakes specialist procedural competency assessment.

These initiatives could reduce a number of potential biases, strengthen quality assurance through confirming delivery of the intended intervention and defining intraoperative protocol deviations (110, 225). Objective surgical skill data also may assist the evaluation and interpretation of trial results as well as detection and quantification of performance bias and its impact. These issues are not confined to trials. The need for improvement in routine practice persists and tool data could assist established programs such as the US national accreditation program for rectal cancer with accreditation and benchmarking (102, 226). The L-TMEpt provides reliable and objective surgeon specific data that could inform targeted training and quality improvement methods aiming to enhance outcomes.

Limitations and areas for future study

Whilst the aims of this chapter were successfully met it should be reviewed in light of some limitations. The nature of the video series meant it was not possible to utilise the full L-TMEpt instrument in all cases and a number of laparoscopic ALaCaRT cases were never recorded. However, TME dissection, as would be expected to account for specimen quality, was uniformly analysed. It should be remembered that the L-TMEpt is restricted to the assessment of technical performance and does not consider non-technical and human factors such as team composition, the operating theatre environment, distraction and fatigue all of which may impact performance (220, 227, 228).

Importantly the study design and review of completed RCTs means no comment regarding the predictive validity of the L-TMEpt can be made. Longitudinal data is not held and would be of great interest to investigate whether observed individual performance levels alter and to what degree. In the only potential comparable study, the Michigan bariatric group performed blinded analysis on a second case from each participating surgeon with very strong correlation observed (r=0.85) (130). It is unclear if specialist surgical performance plateaus upon completion of proficiency curves or improvement remains possible. Returning to the sport analogy, elite athletes continue to receive training at every step of the career even when a great success is achieved. It is highly doubtful that continuing medical education requirements beyond completion of surgical training are comparable.

It may not be possible to eliminate variability in surgical practice, but recognition and measurement are important early steps towards quality improvement. Until performance is first quantified and areas of need defined, targeted improvement is not possible. Non-targeted improvement efforts risk being ineffective if they focus on the wrong factor or area of proficient performance. This research has successfully identified that surgical skill varies and it is hoped that improvement is possible. However, this cannot be assumed as both trials and this chapter did not feedback data to individual surgeons or attempt to provide a targeted intervention such as coaching, mentoring or highlight areas shown to fall below the peer group average. These are an essential requirement before this QA strategy can be considered for wider implementation within clinical settings.

This research arbitrarily divided surgeons and cases into quartiles. Although clinically significant results were seen this strategy is inappropriate if the intraoperative QA is to be widely applied. It was not possible to define a performance threshold associated with acceptable outcomes. This requires ongoing study as a robust figure would be needed if the studied tools were used in high stakes settings such as credentialing and practice rights. In addition to the appropriate supportive statistically analyses such as AUROC, this work would require patient and public involvement and input from regulatory bodies given the implications of low scores for patients, surgeons and healthcare providers.

Additionally, an agreed performance threshold would greatly assist surgical study design as demonstration of sufficient skill could provide further QA within surgical interventional studies.

The analyses were based on video-based observation resulting in a cleaner dataset and standardised analysis setting and method. As all surgeons were aware of video recording there a is potential risk of the Hawthorne effect which has yet to be investigated in surgical practice. Such a study would be tricky to design and perform given the need to obtain patient consent and capture case video whilst ensuring the operating surgeon and ideally entire theatre team remain unaware.

Application of the L-TMEpt to routine practice would require direct in-theatre assessment of performance, and it is an understudied question if results are comparable with those obtained in the experimental setting. In a cadaver based study, direct observation was seen to hold advantages over objective assessment of laparoscopic cholecystectomy video based performance at the trainee level (229). It is not known if the assessor would be influenced by interactions with the surgical team and theatre environment. Acceptability to practicing surgeons is another consideration before wider application of the L-TMEpt can be undertaken. Although the tool was designed in conjunction with a large number of global laparoscopic TME experts and piloted with a number of UK experts, it is unknown if intraoperative performance assessment would be tolerated by specialist surgeons. Particularly if it were linked to potential controversies such as practice restriction or demonstration of an appropriate level of skill in a similar fashion to the mandatory simulator assessments required in industries such as commercial aviation. A qualitative exploration of surgical culture with involvement of all stakeholders including regulatory bodies is needed.

The L-TMEpt is designed to facilitate categorical qualitative appraisal of skill areas within a series of tasks with the assumption that each item is of equal importance (129). Further work is required to define the relative importance of each on overall performance and outcome. If quantitative analysis were applied for determining future pass/fail thresholds, it would prove challenging to investigate

whether linear scale differences i.e. between score 3 and 4 is the same as that between 1 and 2 (129). This would need to be addressed before high stakes assessment use is contemplated.

All cases were performed by credentialed specialist surgeons meaning the wider applicability of the L-TMEpt outside this setting is presently unknown. The tool holds theoretical applications in formative training and summative demonstration of competency settings, with the potential to structure feedback, track progress and identify specific areas for improvement. Dedicated study is underway.

Finally, in these three chapters patients were enrolled to their respective studies on the basis of undergoing surgery with curative intent. Although this was an explicit inclusion criteria in both trials, we did not perform a central case note, MDT record or radiology review to ensure that in all operations it was possible and expected to obtain mesorectal fascial plane surgery, clear margins and R0 resection. Such quality assurance would have further strengthened this study.

5.5 Conclusion

The application of intraoperative quality assurance objectively and reliably measured surgical performance in complex cancer interventions within two laparoscopic TME randomised trials. Surgical skill varies and is shown to be a key determinate of outcomes after cancer surgery. Intraoperative QA tools identified high and low performing surgeons with significant associated impacts on perioperative, histopathology and morbidity outcome data. A pragmatic, clinically applicable intraoperative reporting system could assist the uptake of QA in minimally invasive surgical practice.

Chapter 6 – Development and validation of an intraoperative adverse event and near misses reporting classification

The previous chapters have shown that the application of intraoperative quality assurance and the resulting data adds to our understanding of intervention delivery and the associated clinical outcomes. During the extensive video analyses performed further opportunities for intraoperative QA were identified. A pragmatic and simpler system could further enhance the available QA strategies and assist the routine adoption of QA in surgery. Here a novel intraoperative adverse event severity classification is developed and validated using the video series analysed in this thesis.

This work was shaped by the expertise of the following members of the EAES executive and committee boards who participated in the surveys. I gratefully acknowledge the expert input from: Abu-Hilal M (Southampton, UK), Agresta F (Adria, Italy), Antoniou SA (Athens, Greece), Arezzo A (Torino, Italy), Balagúe C (Barcelona, Spain), Boni L (Milan, Italy), Bouvy N (Maastricht, Netherlands), Carus T (Hamburg, Germany), Edwin B (Oslo, Norway), Diana M (Strasbourg, France), Faria G (Porto, Portugal), Ignjatovic D (Oslo, Norway), de Manzini N (Trieste, Italy), Margallo FM (Cáceres, Spain), Martinek L (Prague, Czech Rep.), Matveev N (Moscow, Russia), Mintz Y (Jerusalem, Israel), Nakajima K (Osaka, Japan), Popa DE (Bucharest, Romania), Schijven PJ (Amsterdam, Netherlands), Sedman P (Hull, UK), Yiannakopoulou E (Athens, Greece). Special thanks go to Dr Stella Marvoveli (Imperial College London) for her expert assistance in refining the wording and structure of the questionnaire.

Mr Godwin Dennison (Consultant surgeon, Yeovil District Hospital) kindly served as the second assessor for the inter-rater reliability analysis.

6.1 Copyright statement

The original figures 6.1 and 6.2 and table 6.1 displayed here were subsequently published (230). They are presented here with written permission (appendix 1).

6.2 Introduction

Intraoperative quality assurance has been shown to be associated with clinical outcomes after laparoscopic TME surgery. The measurement of surgeon performance complements routinely collated datasets in both clinical and research settings. Efforts to increase the usability of intraoperative QA are now justified to assist the uptake and utilisation of intraoperative QA particularly within clinical practice. The two QA methods studied in this thesis could presently be considered as research tools. Simpler and pragmatic clinically applicable systems could act as bridge towards successful widespread use of intraoperative QA within clinical settings. As shown in chapter two, no such system has been reported.

Whilst undertaking the approximately 1800 hours of video analysis required for the objective performance assessments described in the previous three chapters it become apparent that intraoperative QA could be further enhanced. At an early stage it was apparent the studied intraoperative error event reporting mechanisms used could be considered insufficient as they did not necessarily capture the differing consequences of identified events. Both methods studied in this thesis could be considered

OCHRA is based on the identification and categorisation of error events. This raises two concerns: the assessor is required to set and maintain their error threshold and the binary reporting system of acceptable or error does not reflect the reality and the range of impact (if any) on the patient, procedure, surgeon and healthcare resources. For example, intraoperative bleeding ranges in both volume and consequence from the irrelevant where the procedure continues safely without the need for any corrective step or subsequent alteration in care. Whereas the other end of the

spectrum, a major vessel injury with catastrophic haemorrhage, presents a direct risk to life with the need for immediate extensive correction and significant detrimental effects to the patient and healthcare provider with the need for additional resources such as blood products and critical care utilisation. Presently OCHRA would consider both these events similarly as one error. Differing consequence category codes can act as a discriminator but could still mask significant clinically relevant heterogeneity (section 3.7). Therefore, it could be argued the adopted objective assessment tools and potentially intraoperative reporting of MAS in general would be strengthened through the application of a grading classification designed to capture the impact of identified error events.

There are a number of standardised and clinically applicable surgical classification systems widely used in other areas of peri-operative care. Examples include the histopathological assessment of resected cancer specimens (14), the Clavien-Dindo classification of post-operative morbidity (183), the American Society of Anaesthesiologists physical status classification (231) and the Mallampati classification to predict difficulty in endotracheal intubation (232). Such systems can reduce subjectivity, variability and under-reporting. These examples all aid the reporting of their respective areas and their widespread use assists direct patient care, handover, and data gathering for clinical governance and research use through score comparison. It is noteworthy that there remains no accepted and clinically applicable measure for assessing the delivery of laparoscopic surgery or reporting of intraoperative events. This has led to uncertainty and variations in practice with concerns regarding the underreporting of the intraoperative period raised (166, 233, 234).

Overall, a pragmatic and simpler system could be more clinically applicable and generalisable with the potential to facilitate greater uptake in the application of intraoperative quality assurance techniques to routine MAS surgery.

Aim

To design and pilot a MAS intraoperative quality assurance adverse event classification applicable to routine clinical practice and acceptable to expert surgeons.

Chapter objectives

- Create a generalisable classification for the reporting of intraoperative MAS errors
- To establish a consensus amongst expert surgeons on:
    - The desired characteristics of such a system
    - The acceptability of the reporting classification
    - Potential uses in their practice
    - Potential pitfalls and barriers to uptake
- Pilot the classification on intraoperative case videos
- Explore the clinical validity of the classification
- Report the reliability of the newly developed system
- Report the frequency, nature and impact of non-consequential "near miss" errors

6.3 Methods

Conception and expert surgeon questionnaire design

A steering group of surgical researchers with experience of objective assessment, MAS and tool development was formed to undertake this work consisting of my supervisors, Mr John Conti (Consultant surgeon, Portsmouth), Mr Jake Foster (Imperial College London), Prof Jaap Bonjer (Past president of the EAES and Professor of Surgery, Amsterdam) and myself.

From the outset a structured mixed methodology approach was adopted as it was recognised the work required multiple threads for success. Involvement of expert surgeons was considered essential in both the design and evaluation phases as well as maximising awareness and potential uptake of any classification. A bespoke questionnaire was created using the SurveyMonkey® web platform (San Mateo, CA, USA). Firstly, current intraoperative reporting practices and utilisation of existing assessment tools or methods was explored. Surgeons were then directly asked if they knew of any ideal system for intraoperative reporting. To investigate the need for any classification, surgeons used a five-point Likert scale to report the possible impact upon training, post-operative care, research (including MAS trials), quality assurance and surgeon credentialing. Additional free

text responses were permitted to allow free expression of ideas and implications. The survey underwent refinement by a psychologist with experience in questionnaire design and can be viewed here: https://www.surveymonkey.co.uk/r/JRV555L. After internal piloting and final agreement on all questions and formats, the survey link was electronically sent to the EAES executive board and members of the research, technology and education and training committees in March 2017. This approach was felt likely to obtain the opinions of a broad range of expert surgeons and an acceptable response rate. To take EAES office, practising surgeons must be board certified with expertise in laparoscopic clinical practice and selected by membership vote.

Ten system characteristics were then displayed with respondents asked to order all ten from most desirable to least important features of an ideal intraoperative reporting classification. Additionally, they were asked to select their single most important option. The list comprised: completion time, simplicity, acceptability, reproducibility, logical, useful, comprehensive, clinically applicable, usable and analysis time.

Classification development

The initial creation and refinement of the grading system utilised the then available 44 unedited laparoscopic total mesorectal excision (TME) case videos from 2D/3D multicentre randomised controlled trial described in chapter 4. Approval for the adhoc use of the source videos was granted by the chief investigator, sponsor and 2D/3D trial management group. At the time of this work the main trial was ongoing meaning this sub-study had no access any trial data including operative, histopathological or patient data which remained blinded with no interim analyses performed. At least three months following initial analysis of the deidentified 2D videos, all previously identified OCHRA error events were sequentially revisited with the preceding and following 30 seconds which were viewed using Adobe Premiere Pro (v6.0, Abode, San Jose, CA, USA).

During this undertaking it was observed that immediately following errors there appeared to be two primary responses: procedure continues (minor error, no correction needed or missed by surgeon) or error requires rectification (corrective step or alteration to procedure). As observation continued,

I observed a wide variation in the nature of corrective steps. Many were minor such as application of diathermy to a bleeding point whereas other more serious events were infrequent (splenic capsule tear, full thickness posterior vaginal wall injury) but required more extensive repairs and their impact was likely to extend outside of theatre through alteration of post-operative care and impact on patient recovery. A pyramidal structure was the logical outcome from this stage. Non-consequential error (defined as no damage or corrective action required, procedure continues unchanged) was adopted as the bottom layer of the pyramid and based on ongoing video review four grades of consequence (minor, major, life threatening and fatal) were selected to make a draft five-point scale. This sufficiently captured the impact of events without risking classification errors from higher number of possible options. Remaining videos were used for continuous ongoing refinement. Where errors could not be easily classified, they were reviewed by the committee and agreement or alteration reached. When all observed errors could be classified the initial classification was set.

Second questionnaire phase

After completion of the initial questions, the draft classification was made available to the expert surgeons. Requested feedback focussed on their agreement with the tool and its usability, acceptability, clinical applicability, face and content validity and potential uses within their practice all using five-point Likert scales. Surgeons were asked to rate the classification against the same ten ideal characteristics previously displayed this time using a ten point scale. Participants were then encouraged to further shape development through free text comment on any aspects of the system. This data in the ongoing refinement of the classification which continued to be developed in response to newly available TME case videos. Once the present iteration was agreed by all members of the steering committee all OCHRA error events were revisited once again with an additional categorical severity code added to each event.

Expert feedback was limited to questionnaire based input as there was insufficient funding to permit consensus workshops and insufficient timetable space or room availability to facilitate face to face meeting during planned EAES events. Additionally, it was anticipated that expert surgeon

attendance might be low due to competing pressures, multinational geographical considerations and perceived level of interest. The questionnaire was administered in one phase to maximise uptake and overcome the traditionally very low response rate to email survey invitations (235, 236).

Clinical validity

At the time of the classification development work the 2D/3D trial was recruiting with all clinical and outcome data remaining blinded. Therefore, clinical validity analyses were performed after completion of the trial. Two datasets were studied:

1) Entire unedited laparoscopic TME cases from the 2D3D trial (n=77).
2) The combined ALaCaRT and 2D3D series as described in section 5.2 (n=176).

Two analyses were performed:

1) Individual cases were labelled by highest identified EAES grade event present.
2) The total number of non-consequential events per case.

As defined by the classification, a serious event would be expected to have a significant clinical impact with major events potentially holding a disproportionate influence over patient outcomes. Cases were labelled by the highest grade event identified by OCHRA assessment. In an attempt to move away from video review, labelling the case by the highest observed severity could be applied to routine practice and it could be hypothesised that such an event has a dominant effect upon the post-operative course of the patient. This score was considered as an ordinal rank to group cases.

A second analysis was based on the total number of non-consequential error events. In the development work, the majority of observed adverse events were noted as grade 1 and as there is no prior research on non-consequential events in surgery an exploration of their nature, distribution and link with post-operative events is warranted. This sum was managed as a continuous scale variable during analysis.

Clinical validity endpoints studied were plane of mesorectal dissection, number of 30 day complications, highest observed Clavien-Dindo classification event per case and length of stay. Although there were few within the trial, given their importance to patients and outcomes anastomotic leak and unplanned reoperations were also studied (237). In both analyses, lower scores are considered to represent higher quality operations.

Reliability assessments

Test-retest reliability was explored by blinded reassessment of all 2D3D case videos for all previously identified adverse events at least one year after first analysis. Inter-rater reliability was performed by a consultant surgeon acting as second assessor. They were independent from the trial and blinded to all procedure and outcome data. After receiving training on the classification and its purpose they were granted sufficient time to review the published manuscript before practicing applying the classification to errors events present in case videos that had previously undergone OCHRA review as part of a separate completed TME RCT (140).

Statistical analysis

Quantitative data were reported using descriptives and analysed using SPSS® (version 24, SPSS Inc., IL, USA). Means and standard deviations are reported unless otherwise specified. Qualitative data were thematically analysed to identify commonly reported factors. Response data graphs were generated by the inbuilt Survey Monkey® website tools (San Mateo, CA, USA). After exploration for normality with the Shapiro–Wilk test and detrended Q–Q plots, data was compared with T-Testing to compare means. Inter-rater and test retest reliability was explored using Cohen's kappa (κ) statistic with 95% confidence intervals calculated using the standard error. Interpretation of correlation strength was quantified as excellent ≥0.75, good 0.74-0.6, fair 0.59-0.4 and poor ≤0.39. $P<0.05$ was considered significant.

6.4 Results

Expert surgeon participants

Thirty-four members (of 49, 69%) of the EAES executive and committee boards completed the survey. Respondents were based across 14 nations with significant MAS expertise (median 20 years (range 9–30) experience and median 3500 laparoscopic cases (500–10,000)). 85% self-reported their role as professor of surgery and/or chief of surgery at their institution.

Survey responses

Twenty-three (68%) stated they routinely reported intraoperative adverse events, primarily on operative notes (86%) or informally with colleagues (45%). Some 86% stated they did not use any structured or previously reported system. Subjective review of case video was the most commonly reported strategy for reviewing their procedures, but this was not commonplace (21% of respondents). Amongst the expert surgeons stating they do not presently report on table events, 70% said their main barrier was the lack of an appropriate system. Overall, 96% felt no ideal system presently exists for this purpose. Respondents felt a reporting classification could positively impact post-operative care (Likert scale, mean 4.42), training (4.5), research and minimally invasive surgical trials (4.19), quality assurance (4.62) and accreditation of surgeons (3.92). By far the most commonly reported free text theme was concerns about the term “error” and potential medicolegal implications from acknowledging errors enacted during their operations. 60% favoured “intraoperative adverse event reporting” as the most appropriate title for a classification.

Ideal characteristics

Simplicity (7.6), reproducibility (7.0) and clinical applicability (6.7) were regarded as the most important considerations with analysis time (3.7), acceptability (4.5), time to complete (4.5) and comprehensive (4.6) the least supported. 35% selected clinically applicable as their single most important consideration of an ideal intraoperative reporting classification.

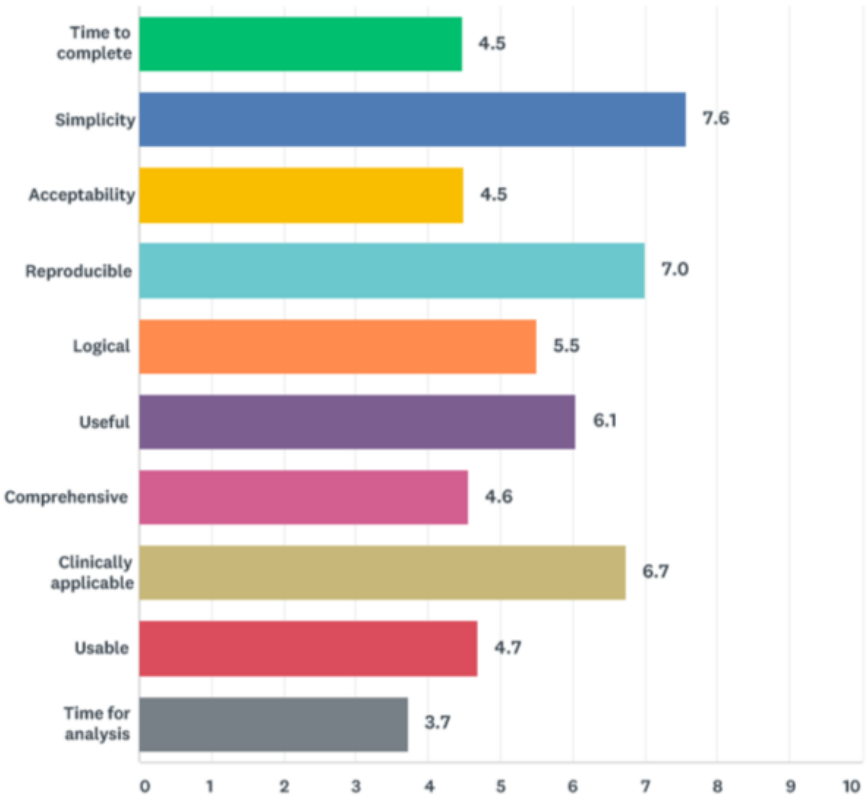


Figure 6.1: Expert reported ideal characteristics. Each participant ranked all ten which were awarded a score between 10 (favourite) and 1 (least). Figures represent means.

This data was used to refine the system. The steering committee felt simplicity had been met but reproducibility was not. There was debate regarding possible subjective reporting and appropriate boundaries between the grades. Where a surgeon self-reports their performance there is potential for downplaying and under reporting and their scores may be discordant from those from an external assessor. A reduction in subjectively may further improve reproducibility and aid clinical applicability.

The Clavien-Dindo classification of post-operative morbidity was reviewed (182, 183). This structured five-point hierarchical reporting classification is widely used and acceptable to surgeons despite quantifying and classifying morbidity following their interventions. It is based on the therapeutic requirements to remedy each event which presents a number of strengths. This method could be expected to reduce subjectivity and under-reporting and is more likely to reflect the true

impact upon the patient and healthcare provider. The classification also assists clinical governance and research by categorising the complex post-operative period which aids comparison and benchmarking. This strategy has been shown to be simple, objective and reproducible across different healthcare systems and cultures and acceptable to a wide variety of surgeons (181).

The hierarchical layers of the system were rebranded based upon the therapeutic consequences the observed adverse event. Grade 1 remained no corrective action needed but grades 2-4 could be better differentiated through minor corrective action with no post-operative impact to major correction and/or change in care and life-threatening implication. This approach was strongly supported by the committee and incorporated in the grading system.

Intraoperative adverse event classification

The present iteration of the classification is displayed with the observed adverse event distribution from the 2D/3D trial videos.

Grade	Intraoperative adverse event classification	Percentage of adverse events identified in pilot	Examples in laparoscopic colorectal surgery
1	Minor error, no damage or corrective action required	60.1%	Peritoneal tear; tissue avulsion from excess traction
2	Minor consequential error requiring corrective action but no change in postoperative care	37.1%	Bleeding requiring a corrective step (swab/diathermy), wrong plane dissection - corrected
3	Consequential error requiring major corrective action and/or change in postoperative pathway	2.4%	Splenic capsule tear; re-do anastomosis; enterotomy, unplanned stoma
4	Life-threatening complication that requires major or immediate corrective action and led to a significant alteration of the post-operative pathway which may include re-operation or intensive care admission	0.1%	Visceral injury; major haemorrhage; any complication necessitating unplanned critical care admission
5	Major consequential error resulting in death	0	Major vessel injury, air embolus

Table 6.1: The EAES intraoperative adverse event classification with pilot data distribution.

After review by the expert surgeon participants, 84% agreed or strongly agreed with the proposed classification in its present form (Likert scale mean 4.04).

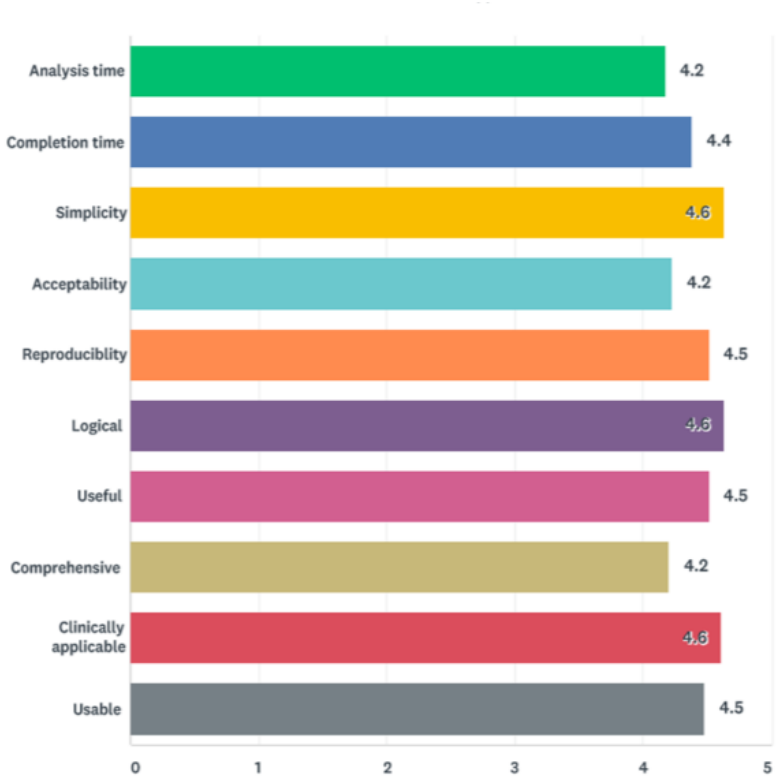


Figure 6.2: Experts graded the classification for each for the ten characteristics using a five point Likert scale. Figures represent means.

The experts were asked to assess the classification against the same ten ideal characteristics previously listed using a five point Likert scale. All ten were deemed acceptable including the aims of meeting clinical applicability, simplicity and reproducibility (all means 4.2-4.6). The majority reported that the system would be useful in their clinical practice (92%), training activities (76%), clinical trial participation (64%) and quality assurance work (56%).

44% felt the system was ready for clinical use in its current format. 44% sought minor grade definition amendments with responses equally split between adding and removing subcategories. This approach was not supported by the steering committee.

Thematic analysis identified support for its simplicity and comparison with other accepted morbidity classifications. The link to consequences of adverse events was strongly supported. Delayed completion was identified with respondents suggesting this would allow correlation with post-

operative events and ensure validity. Some surgeons documented reporting adverse events that hold no consequence may not be a worthwhile exercise and raise possible medicolegal issues.

6.5 Reliability of the EAES classification

Excellent inter-rater reliability was seen (κ=0.86 (95%CI 0.835-0.885), absolute agreement 92.7%, p<0.001). In the 101 events with a discrepancy, 94 differed by one grade. 35 (2.5%) of events were not considered to represent an error by the second examiner.

Blinded test retest reliability assessment was performed on average 12 months after first assessment (range 9-15). Excellent agreement was observed with 93.6% of events received an identical grade (κ=0.887 (95%CI 0.862-0.912), p<0.001). There were seven two grade discrepancies amongst the 88 differing scores (8%). 2% of previous identified errors were not considered as an error at the second viewing.

		Assessor 1					
	EAES grade	1	2	3	4	Total	%
Assessor 2	No error	30	5	0	0	35	2.5%
	1	751	33	2	0	786	57.1%
	2	29	483	2	0	514	37.3%
	3	0	0	39	0	39	2.8%
	4	0	0	0	3	3	0.2%
Total		810	521	43	3	1377	
%		58.8%	37.8%	3.1%	0.2%		

Table 6.2: Inter-rater reliability assessment data.

		Assessment one					
	EAES grade	1	2	3	4	Total	%
Assessment two	No error	23	5	0	0	28	2%
	1	762	32	2	0	796	57.8%
	2	25	484	1	0	510	37%
	3	0	0	40	0	40	2.9%
	4	0	0	0	3	3	0.2%
Total		810	521	43	3	1377	
%		58.8%	37.8%	3.1%	0.2%		

Table 6.3: Test retest reliability data for the EAES classification.

6.7 Clinical validation of the intraoperative adverse events and near miss classification

2D3D cases - highest grade event per case

Labelling the 77 analysed 2D3D trial cases by the highest EAES classification grade recorded during analysis resulted 41 grade 2 (53.3%), 33 grade 3 (42.9%) with three grade 4 (3.9%) procedures. There was no grade 0, 1 or 5 operations within the trial.

		Any 30 day morbidity?		
		No	Yes	Total
Max EAES grade per case	2	20	21 (51%)	41
	3	10	23 (70%)	33
	4	0	3 (100%)	3
Total		30	47	77

Table 6.4: Crosstabulation of complication data with maximum EAES grade data. A significant increase in the proportion of cases that had morbidity events was seen with each increasing grade (p=0.043).

Half of grade 2 cases developed a morbidity event, but this significantly increased in grade 3 and 4 operations (p=0.043). The mean number of complications non-significantly increased between EAES grades 2-4 (1.05±1.3 vs. 1.48±1.3 vs. 2.33±0.6, p=0.145). The highest recorded Clavien-Dindo grade per patient significantly increased between EAES grades 2-4 (median 1 vs. 2 vs. 3, p=0.023). Length of stay also increased (median 6 vs. 7 vs. 61 days, p<0.001). Anastomotic leak and unplanned re-operation were correctly captured by the allocated EAES grade (2.5% vs. 3.3% vs. 100%, p<0.001 and 5% vs. 0% vs. 66%, p<0.001 respectively). Readmission to hospital was not different between the EAES grade 2-4 groups (17% vs. 27% vs. 0%, p=0.373). No clinically applicable pattern is seen with any histopathologically endpoint (data not shown, all p>0.442).

2D3D cases - sum of non-consequential events

Of the 1377 errors identified from the 2D3D trial videos, 809 (58.8%) were labelled EAES grade 1 with a median 10 per case was 10 (IQR 7-13, range 1-28).

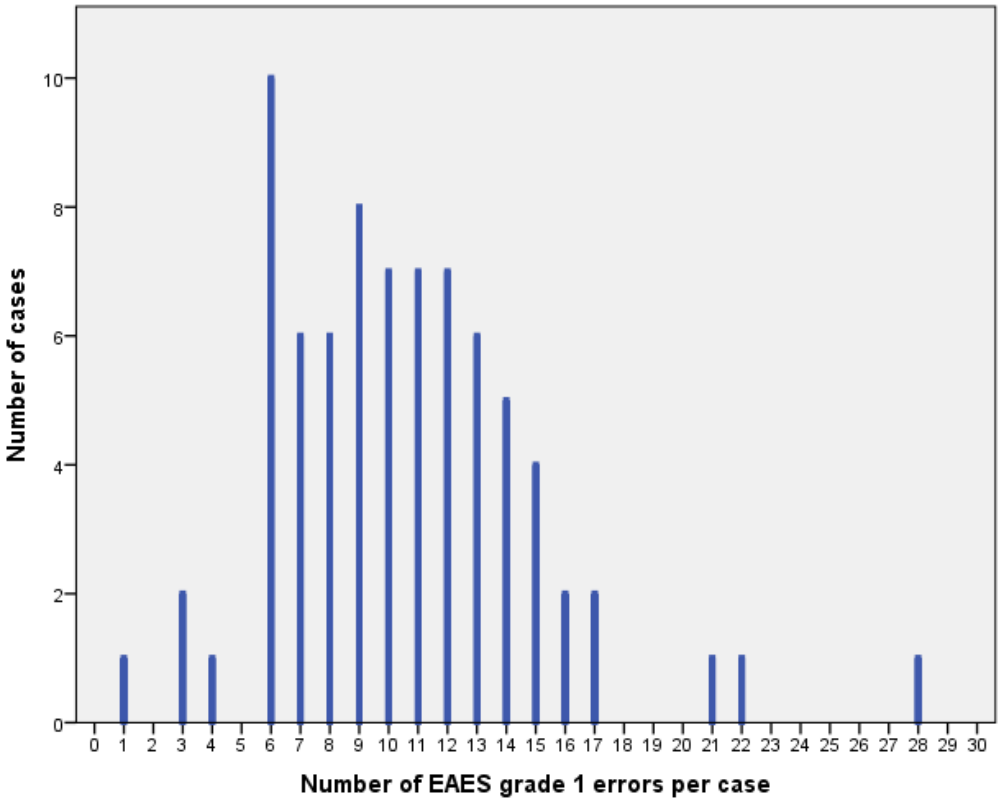


Figure 6.3: Distribution of the number of EAES grade 1 error events per case. All cases contained at least one of these errors.

No differences were seen between mesorectal plane of surgery specimens and intramesorectal or muscularis propria resections (median 10 (7-14) vs. 10 (7-12), p=0.884).

There were significantly more inconsequential intraoperative errors recorded in patients that developed any early morbidity event than those who had an uneventful post-operative stay and 30 day recovery (median 8.5 (IQR 6-12) vs. 11 (9-14), p=0.005). When broken down by number of 30 day morbidity events a significant increase in EAES grade 1 errors is seen (p=0.047).

	Number of 30 day complications											
	0		1		2		3		4		5	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR	Median	IQR
EAES grade 1 errors	8.5	6-12	11	9-13	11	9-14	12	9-16	9	8-9	17	11-22

Table 6.5: A significant increase is seen in the number of apparently inconsequential errors observed and the number of 30 day morbidities recorded (p=0.047).

Weak positive correlation is observed between the sum of EAES grade 1 errors and length of post-operative stay (r^s=0.36, p=0.001) with linear regression confirming the weak association between the variables (y=10+0.04*x, goodness of fit r^2=0.017).

No link between grade 1 error count and anastomotic leak (median 10 vs. 10, p=0.986), re-operation (9 vs. 10, p=0.991) or readmission was seen (9.5 vs. 10, p=0.659).

Combined ALaCaRT and 2D3D pelvic tasks - highest grade event per case

Labelling the 175 ALaCaRT and 2D3D pelvic phases by the highest EAES classification grade recorded during analysis resulted in a wider distribution: no pelvic errors 4 (2.3%), 32 grade 1 (18.3%), 109 grade 2 (62.3%), 26 grade 3 (14.9%) with four grade 4 (2.3%) procedures.

		30-day morbidity			
		None		Any	
		Count	Column %	Count	Column %
Highest EAES pelvic grade event (per patient)	None	4	4.6%	0	0.0%
	1	19	21.3%	13	15.3%
	2	55	61.8%	53	62.4%
	3	10	11.2%	16	18.8%
	4	1	1.1%	3	3.5%

Table 6.6: Crosstabulation showing highest EAES grade present in pelvic surgery and the development of 30-day morbidity. Higher grades were present in the morbidity group (p=0.015).

When the cohort is divided into those that developed any 30 day morbidity (50.9%), there are lower EAES grade cases in the no morbidity group (p=0.015). The distribution is visually comparable suggesting that there are other influential factors such as abdominal phase data, satisfactory remedy of serious error events and post-operative management.

		Highest EAES pelvic error grade per case									
		None		1		2		3		4	
		Count	%	Count	%	Count	%	Count	%	Count	%
Number of 30-day morbidity events	0	5	100.0%	19	59.4%	55	50.9%	10	38.5%	1	25.0%
	1	0	0.0%	8	25.0%	21	19.4%	7	26.9%	0	0.0%
	2	0	0.0%	3	9.4%	19	17.6%	2	7.7%	2	50.0%
	3	0	0.0%	1	3.1%	6	5.6%	6	23.1%	1	25.0%
	4	0	0.0%	1	3.1%	5	4.6%	1	3.8%	0	0.0%
	5	0	0.0%	0	0.0%	2	1.9%	0	0.0%	0	0.0%
Highest Clavien-Dindo event per patient	0	5	100.0%	19	59.4%	55	50.9%	10	38.5%	1	25.0%
	1	0	0.0%	3	9.4%	8	7.4%	4	15.4%	0	0.0%
	2	0	0.0%	5	15.6%	31	28.7%	9	34.6%	0	0.0%
	3	0	0.0%	5	15.6%	9	8.3%	2	7.7%	3	75.0%
	4	0	0.0%	0	0.0%	5	4.6%	1	3.8%	0	0.0%

Table 6.7: Cross tabulation showing EAES grade distribution and the number of morbidity events and highest Clavien-Dindo grade recorded per patient.

EAES grade labels were associated with the number of 30-day morbidities (p=0.046) and the highest Clavien-Dindo grade experienced but the latter did not reach statistical significance (p=0.06).

Median length of stay increased in the presence of any error (5 days (3-5) vs. 8 (7-12), p=0.023). All patients with an anastomotic leak (p<0.001), reoperation (p=0.001) or hospital readmission were correctly labelled as EAES>2 but readmission was not associated with the highest EAES grade label (p=0.399). Similar to the 2D3D dataset, the EAES label was not associated with any histopathology endpoint (all p≥0.144).

Combined ALaCART and 2D3D pelvic tasks - sum of non-consequential events

A total of 1113 errors events were captured during TME dissection, bowel transection and anastomotic components of the procedures. All error events were successfully categorized with the EAES classification (table 6.1). 698 near miss events (62.7%) were identified in the pelvic task phases with significant variation in counts observed between included cases (median 3, interquartile range 2-5, mode 3, range 0-15, p<0.001) (238). Only 11 cases (6.3%) were error free.

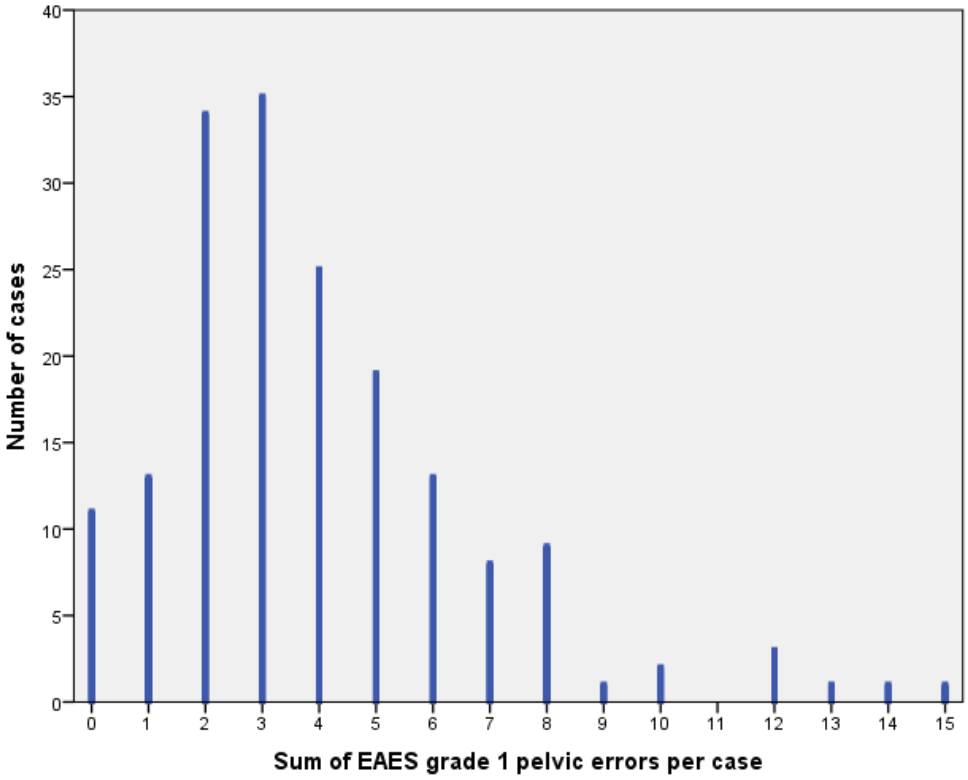


Figure 6.4: The distribution of EAES grade 1 near misses in the pelvic task phases.

There was a stepwise decrease in event frequency in each additional EAES grade (EAES2 379 (34%), EAES3 33 (3%), EAES4 3 (0.3%), table 2). One major intraoperative event was observed for every 19.4 near misses.

EAES Grade	Intraoperative adverse event classification	Sum and percentage of adverse events identified	Cumulative percentage	Number of adverse events per case (median, interquartile range, range)
1	Minor error, no damage or corrective action required	698 (62.7%)	62.7%	3, 2-5, 0-15
2	Minor consequential error requiring corrective action but no change in postoperative care	379 (34%)	96.7%	1, 0-2, 0-10
3	Consequential error requiring major corrective action and/or change in postoperative pathway	33 (3%)	99.7%	0, 0, 0-2
4	Life-threatening complication that requires major or immediate corrective action which led to a significant alteration of the post-operative pathway which may include re-operation or intensive care admission	3 (0.3%)	100%	0, 0, 0-1
5	Major consequential error resulting in death	0	100%	
Total		1113	100%	5, 3-8, 0-31

Table 6.8: EAES classification with pelvic error data from the combined trial dataset. The distribution follows that of the pilot data shown in table 6.1

Cases containing any major intraoperative event (EAES3 and/or EAES4, n=45, 25.7%) contained significantly more errors and near misses (8 (6-12) vs. 5 (3-7); 5 (3-7) vs. 3 (2-5) respectively, both $p<0.001$). 85 (48.6%) patients developed any morbidity within 30 days with significantly higher numbers of EAES1 errors seen in these cases (4 (3-6) vs. 3 (2-4), $p<0.001$). Near misses were significantly associated with the number of 30-day morbidity events as well as the associated

Clavien-Dindo severity encountered (table 3). Higher number of near miss errors were observed in patients that underwent unplanned re-operation or developed an anastomotic leak, but this was not statistically significant (n=10, 5 (2-5) vs 3 (3-7), p=0.172 and n=8, 5 (2-5) vs 3 (3-7), p=0.165 respectively). Near misses were not associated with length of hospital stay (r^s=0.174, p=0.283).

		Total number of pelvic EAES grade 1 intraoperative adverse events			
		Count	Median	Inter-quartile range	p
Any early complication?	None	90	3	2-4	**<0.001**
	Yes	85	4	3-6	
Number of early complications per patient	None	90	3	2-4	**0.002**
	1	36	4	2-5	
	2	26	5	3-7	
	3	14	4	2-8	
	4	7	5	2-6	
	5	2	9	7-10	
Highest Clavien-Dindo grade experienced per patient	None	90	3	2-4	**0.003**
	1	15	4	3-7	
	2	45	4	3-7	
	3	19	3	2-5	
	4	6	5	3-6	
Re-operation	No	165	3	2-5	0.197
	Yes	10	5	3-7	
Anastomotic leak	No	167	3	2-5	0.165
	Yes	8	5	3-7	

Table 6.9: Median number of pelvic near misses associated with defined short term patient outcomes.

As would be expected given the design and purpose of the classification no association between near misses and any TME specific histopathological outcome was seen.

		Total number of pelvic EAES grade 1 intraoperative adverse events			
		Count	Median	IQR	p
TME Plane of surgery	Mesorectal Fascia	141	3	2-5	0.761
	Intramesorectal	17	4	2-5	
	Muscularis Propria	9	5	3-5	
Mesorectal Fascia Plane	No	26	4	2-5	0.577
	Yes	141	3	2-5	
Circumferential margin status	Clear	158	3	2-5	0.426
	Involved (<1mm)	9	3	2-4	
Distal margin status	Clear	173	3	2-5	0.572
	Involved (<1mm)	2	5	3-6	
ALaCaRT Composite Endpoint	Unsuccessful	31 (17.7%)	3	2-5	0.873
	Successful	144 (82.3%)	3	2-5	
pT	Benign	10	3	2-4	0.786
	1	18	3	2-5	
	2	54	3	2-5	
	3	81	3	2-5	
	4	11	4	2-6	
pN	0	116	3	2-5	0.727
	1	40	3	2-5	
	2	19	4	2-6	
R	0	166	3	2-5	0.72
	1	9	3	3-5	
Stage	Benign	2	5	5-5	0.49
	1	49	3	2-6	
	2	44	4	3-6	
	3	76	3	2-4	
	4	4	3	2-5	

Table 6.10: Pelvic near miss data and pathological outcomes. No associations are seen.

6.7 Discussion

In an effort to assist the routine application of intraoperative QA this chapter describes the conception, design, pilot, validation and reliability exploration for a novel classification of adverse events in laparoscopic surgery. Shaped by a multinational group of expert laparoscopic surgeons who desired a simple, reproducible but most importantly a clinically applicable system, the classification utilised the principle of therapeutic consequences. Despite being used to classify morbidity and mortality events, the Clavien-Dindo classification is generally accepted by surgeons and widely applied in routine practice and research settings (181-183). Using therapeutic steps could be expected to reduce subjectivity and more accurately reflect the impact of the event on the patient, procedure and potentially the healthcare organisation. The presented classification appears to meet the stated needs and requirements for such a system and was adopted by the EAES.

Reviewing nearly 1400 errors from 77 complex procedures encompassing a broad range of situations, inter-rater and test-retest reliability is established with extremely high levels of agreement. The classification is also shown to be associated with early morbidity outcomes. Clinically validity is an essential requirement for intraoperative reporting particularly as specialist surgical performance assessment may come with significant connotations if used in high stakes scenarios. Dual analyses were performed on the available RCT datasets. Overall, the data supports the principle that the intraoperative period is a key determinate of morbidity. The nature and strength of associations are strong and concordant across the exploratory analyses but as would be expected does not form the sole explanation for subsequent events. There are innumerable patient and care factors that could influence morbidity outcomes and given the large number of centres and surgeons it was beyond our ability to control for or record them. Nevertheless, it is evident that intraoperative performance, again shown to vary widely, is a major contributor to post-operative events. It is therefore surprising that so many surgical quality improvement programmes focus on

pre and post-operative phases of care pathways over the intervention (147, 239). The findings throughout this thesis support the refocussing of these efforts towards the operating theatre.

Whilst these results can only be considered for laparoscopic TME surgery and external validity is yet to be explored. The wording of each grade is broad and successfully allowed grades to be applied to a wide variety of intraoperative events and situations. Although the classification is yet to be applied outside TME surgery, from the outset was designed to be generalisable across the breadth of MAS practice. This system was not associated with TME specific histopathology outcomes. This is not surprising as it was not designed for any procedure specific or non-morbidity outcome. Application of the classification to other MAS procedures is currently being performed to explore the generalisability of the classification and our findings.

The classification has a number of potential intraoperative quality assurance applications. Within research settings it could assist surgical intervention studies including RCTs through assessing the number and remedy required following adverse or unintended events. This would be expected to utilise case video review and where performed, the classification directly complements the categorical data obtained by OCHRA analysis. As outlined by the expert surgeon respondents the system could be applied to routine practice. Recording a provisional highest adverse grade on operative notes could assist handover of care and early post-operative management decisions. This would then be revisited and finalised at time of discharge or follow up. An observational multi-surgeon observational study focussing on clinical validity is now required before the classification is considered for further use. Whilst correlation with video review data would be needed to validate this approach, in the long term this strategy could avoid labour and time intensive video analysis.

A systematic review identified that a number of groups have previously attempted to classify intraoperative technical errors by severity (221). However, the authors criticised the identified definitions as non-standardised, subjective and arbitrary. 81% of included studies were based on laparoscopic cholecystectomies alone and were seen to mainly use just two groups resulting in the

inappropriate combination of non-comparable events. The authors concluded that the approach adopted here of considering events as consequential and non-consequential with corrective action requirements formed a better tactic (221).

Near miss (non-consequential) events

Technical error events are a common occurrence within specialist performed laparoscopic cases even those with successful outcomes and the absence of complications (132, 188, 240). A particular strength of this work is that to our knowledge this video based analysis represents one of the first reported methods to identify non-consequential events that would typically not be acknowledged or captured by any other reporting mechanism or within routine practice. The majority of expert responses confirmed there was an unmet need for such a system as they did not report their intraoperative adverse events for this reason. This mirrored results from a UK surgeon survey (166).

Our classification could help surgery to follow the longstanding approach taken by high risk industries particularly those that involve the interaction of humans with complex systems (164). Commercial aviation, offshore oil rigs, motorsport and aspects of military practices are commonly cited examples of industries were near-misses and non-consequential events are considered important learning opportunities that can shape performance improvement. Outcomes from errors are highly variable and subject to many factors including luck. Placing focus on the error event itself is fully justified as without correction, there is a risk of a repeated event where the result could be worse. This principle underpins the industrial safety pyramid first reported by Heinrich in 1931 and updated by Frank Bird Jr in the 1970s (241). Based on analysis of insurance data of nearly two million industrial accidents a pyramidal structure was seen with frequent, mostly nonconsequential unsafe situations with much fewer serious accidents. The ratio of 1 major injury for each 29 minor injuries, 300 no injury accidents and 600 unsafe conditions has been repeatedly reproduced and is widely quoted in safety culture as Heinrich's Law (242).

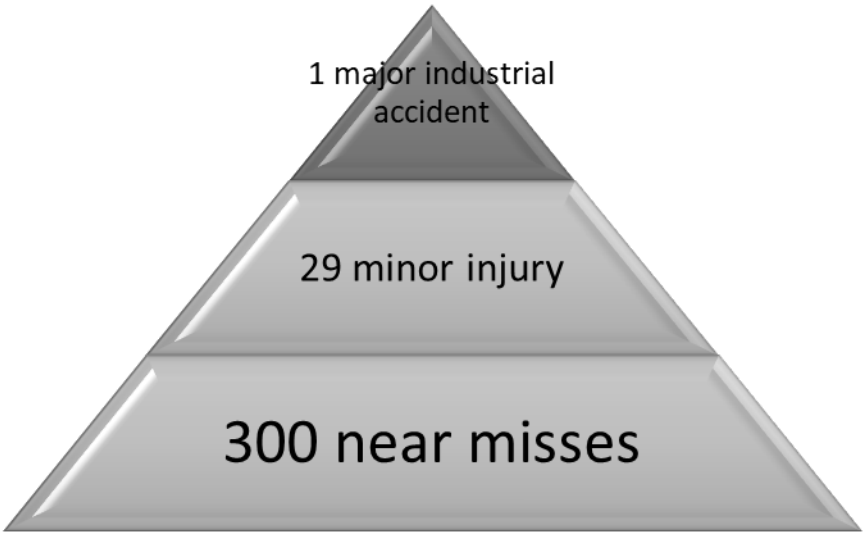


Figure 6.5: Heinrich's law in industrial safety. An additional unsafe conditions tier is sometimes added at the base of the pyramid.

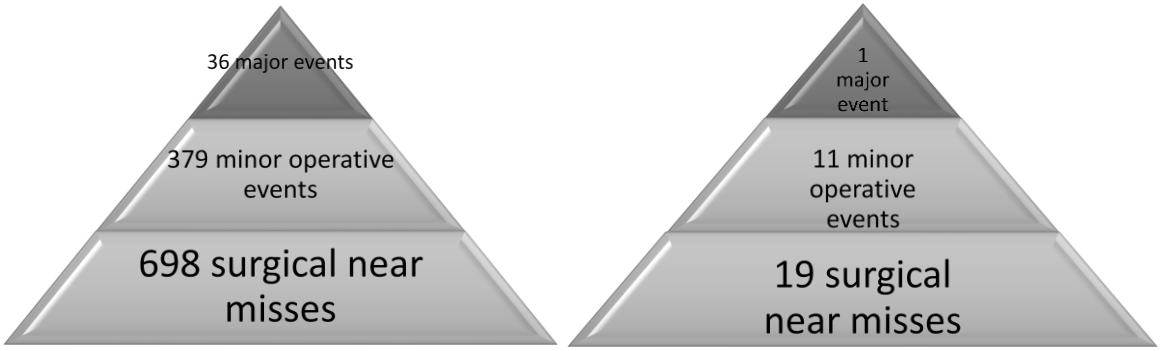


Figure 6.6: The surgical safety pyramid. Total number of events are displayed on the left using the EAES classifications grades. The right hand pyramid displays the ratios for one major event. 19 near misses were associated with each major consequential event presenting an unfavourable comparison with industry.

The observed error frequency from the laparoscopic TME case video is seen to fit this model closely although the surgical error ratio compares unfavourably with industrial data. 1 life threatening event (EAES grade 4) was seen per 14 major errors (EAES 3) with 1 major event occurring for every 12 corrected error (grade 2). Underpinning the pyramid, a grade 2 event was seen for every 1.5 unsafe conditions/non consequential event (EAES 1) (238).

The second conclusion of Heinrich and Bird was that injury occurred as the action of preceding factors which became known as the domino model of accident causation (241). Here an accident is seen a causal chain of events from social environment, the individual, an unsafe situation and then accident and injury. If the key details are known, future events could be prevented through removing a domino and breaking the chain. For example, individuals can be (re)trained and unsafe circumstances removed to avoid an accident (241).

Acknowledgement of near misses forms a fundamental requirement and the first step towards improvement. Of most concern, unlike industry, presently surgeons and healthcare systems could be considered to not be reliably capturing nor learning from these events. Without this data pre-emptive remedy efforts are severely constrained. Root cause analyses, as would be expected following a major event such as unexpected death following surgery, are applied retrospectively after the event. Transition towards a prospective, pre-emptive safety culture could be highly advantageous to patients and healthcare providers but would require a significant cultural change. Without this, widespread improvement is unlikely as without prior awareness or interest in these supposedly non-consequential events little can be learnt.

The EAES classification is demonstrated to be a valid, reliable and clinically applicable system for the assessment of technical errors in laparoscopic TME surgery and represents the first steps along a new route for quality and safety improvement. Being badged by the EAES should assist both the promotion and uptake of the classification and is hoped this represents a new avenue for patient safety through the application of quality assurance to the intraoperative period.

6.8 Limitations and future research

Although the aim of this chapter has been met, the work should be considered in light of a number of limitations. Whilst a broad range of expert surgeons kindly contributed to this study their input was limited to a single administration of an electronic questionnaire. Electronic survey invites have a notoriously poor response rate (235, 236). To combat this, we invited a dedicated group of MAS

experts with a single request. Whilst multiple phases would have constituted a superior method it was strongly felt this would lead to a markedly reduced engagement and uptake. This decision is supported by the high response rate from a range of expert surgeons with significant MAS experience and completion of all survey sections.

Dedicated face to face workshops could have generated invaluable debate and provided a rich context regarding current practices and the use of an intraoperative reporting classification. It was not feasible to perform this due to insufficient funding to reimburse expert travel expenses and provide an appropriate central venue best suited for the 14 participating countries. Furthermore, the expert surgeons were likely to face multiple competing timetable commitments and it was considered this activity may not attract sufficient interest to establish a quorum.

A number of questions remain regarding the potential wider and routine application of the EAES classification which presently prevents its use outside of research settings. This classification was refined using specialist performed laparoscopic TME procedures alone which is acknowledged to be a complex procedure with multiple phases and typically long operating times. It is possible that this influenced the very low levels of grade ≥3 events observed. It is presently unclear if the classification is generalisable to other MAS procedures with different anatomical, ergonomical and procedural steps. The observed adverse event distribution was that of experienced specialists. It is likely that other MAS interventions including those from other subspecialties and those delivered by trainees would result in differing results that would challenge comparative analysis (74, 132). Shorter procedures could be expected to generate fewer adverse events preventing direct comparison with complex resectional procedures. External application to other MAS cases, at other centres, performed by surgeons with differing levels of experience represents an immediate research need.

Although all identified adverse events were successfully categorised a number of theoretical questions remain. Not all error events can be rectified. For example, inadvertent nerve injury meets the error definition applied in this thesis with potential quality of life implications (47, 158).

Additionally, it is possible for surgical teams not to correct an error if they are unaware or miss the event. Despite the extensive video review undertaken here it is unclear on the optimal reporting of such events. This may lead to inaccurate high grade 1 error count as no corrective action took place.

The classification was developed upon the analysis of technical error events. The error definition utilised was chosen following successful use in prior MAS OCHRA studies including TME procedures and allowed classification of all events. There is no consensus on the surgical error definition limiting the ability to compare data with other literature reports in this field (132, 221). Use of an alternative definition could have produced a different error dataset with knock on effects to the classification and validity data. It must be considered that the focus upon technical errors and the psycho-motor domain is likely to represent an oversimplification and it is questionable if it is appropriate to be interpreted as a direct measure of surgical skill. The cognitive and affective domains will influence decision making and may account for the observed technical error. For example, misinterpretation or insufficient knowledge of anatomy may be the root cause of a technical error event. Video analysis alone cannot reliably establish this.

As previously presented, the ALaCaRT video series was focussed on the pelvic surgical phases with very few full length cases captured. 50% of errors occur outside the pelvis and so the inability to assess large parts of the procedure risks missing clinically relevant events with post-operative implications (188). Whilst the presented methods are considered adequate for the exploratory nature of this work and sits alongside the 2D3D full case data analysis, it is not a patient-centred approach. There is overlap as both datasets contained the 2D3D pelvic tasks which is likely to account for much of the concordance observed.

In keeping with the previous chapters, video review presents a number of considerations. All surgeons within the ALaCaRT and 2D3D trials along with all cited video based studies were aware their performance was being captured and would undergo detailed analysis. The Hawthorne effect may apply, where this knowledge consciously or unconsciously influences decision making, technical

performance or surgeon reported data. The Hawthorne effect is well recognised in healthcare studies but has not been explored in surgical practice (243). Such as a study would require careful ethical consideration as surgeons would need to be unaware of recordings for the obtained data to be valid, but video capture would require informed patient consent and ethical approval.

Video review also does not capture nontechnical and human factors that can influence surgical performance. Decision making, fatigue, team composition, experience and working relationships as well as operating room events and distractions and their impact on the procedure cannot be incorporated into this classification (220). This work was performed with external retrospective assessment in the research setting as an extension of OCHRA review which represents a labour-intensive undertaking. It is unclear if live reporting or surgeon self-assessment during or following cases or at the time of discharge would obtain the same results and grade distribution. Particularly for grade 1 near miss events that may continue to be overlooked as non-consequential and therefore unimportant events (132, 164). In the only available report comparing direct observation and video assessment of surgery, weak positive correlation with high internal consistency was seen in a laparoscopic cholecystectomy series (229). Further assessment of this question appears to be required. However, in keeping with the reasons behind its adoption within this thesis, the application of OCHRA generates a rich error dataset with injury mechanisms, locations and consequence information that is unlikely to be obtained with other methods.

The consequence of an adverse event (or omission of an operative step) may not be known until after the procedure. This creates reporting challenges both during the procedure, at the time of video review and ensuring appropriate causation with post-operative or histopathological data. The videos used in this thesis came with matched, dedicated and comprehensive clinical datasets robustly collated from their trial settings. This may not be available in routine practice.

The medico-legal implications of intraoperative adverse event reporting have not been explored. Practical legal frameworks regarding clinical video capture, confidentiality, privacy, storage and

patient records are lacking (244). Regulatory bodies require that professional documentation should be accurate and inclusive, but this is general advice and not designed to cover these situations. Participants from two countries showed particular concern with the implications of admitting adverse events suggesting multiple, jurisdiction specific investigations and reports would be needed.

A far more challenging endeavour would be exploring the lack of transparency in surgery with other high-risk industries. These typically support and mandate reporting of all adverse events including near misses often with confidentiality assured and avoidance of a blame culture. These can be valuable learning opportunities that could prevent future injury and considerable efforts have gone into developing open, supportive cultures that allow honest event reporting (164, 245, 246). Sadly, surgery in general can be argued to presently fall short of this mark.

6.9 Conclusions

A novel, simple, clinically applicable reporting classification of intraoperative laparoscopic adverse events is presented and shown to be valid and reliable. The EAES classification aims to further the application of intraoperative quality assurance and act as a bridge between research and routine practice settings. The system now requires external application.

Chapter 7 – Discussion and Conclusions

Here the work of this thesis is summarised and considered as a whole with reference to the application of QA mechanisms to laparoscopic TME literature and areas of contemporary debate. The clinical implications of the results are discussed together with the identified ongoing research needs.

7.1 Background synopsis and summary of chapters

In chapter one, the rectum, rectal cancer and the surgical challenges this presents are highlighted. The need for high quality surgery that achieves a complete cancer resection with optimal histopathological features is evident and forms a critical prognostic factor for oncological outcomes (41, 76, 77, 101). Pathology based endpoints have been used by many of the completed laparoscopic vs. open TME RCTs that utilised varying methodologies and reported conflicting results (40, 85, 88, 89, 91, 93). MAS advocates continue to perform laparoscopic TME in view of modest benefits to short term patient recovery outcomes but the need for generalised improvement remains in light of the concerns raised (21, 40, 93). Intraoperative QA and surgical performance assessments could represent an avenue for understanding results and assisting quality improvement but is rarely studied in surgical practice.

The literature was systematically reviewed to identify the reported assessment tools that have been applied to laparoscopic TME (161). This revealed surprisingly few reports and the majority represented adapted colonic surgery assessment methods which can be considered inappropriate due to the different procedural steps, anatomical and oncological considerations and the differing colonic and rectal MAS evidence bases. OCHRA and the developmental L-TMEpt were adopted for application within this thesis. The reliability and validity of OCHRA in rectal cancer surgery had previously been establishing and the technique had been successfully applied to specialist level training and as an endpoint in a multicentre rectal cancer RCT (73, 74, 140, 159). The depth provided by case video analysis offers a comprehensive review of intraoperative performance.

The L-TMEpt is a bespoke assessment tool for the specialist level under development to meet the identified need (110, 161, 247). It could complement OCHRA data and importantly its utility offered the potential to serve as a bridge between research and clinical settings but required further study. These requirements were added as further objectives of this research.

These methods were applied to available laparoscopic case video from the ALaCaRT trial (89, 90). This major 24 centre study involving credentialed surgeons investigated the non-inferiority of laparoscopic TME over open laparotomy using a -8% boundary. Non-inferiority could not be established but the reasons behind these results were not otherwise apparent (89, 90). Quality assurance through the application of objective assessment with the two techniques was successfully performed with variation in scores seen. Intraoperative QA was seen to compliment the available RCT data and also have new potential applications in surgical trial methodology.

A bespoke multicentre laparoscopic RCT was designed and prospectively delivered which incorporated intraoperative QA including as the primary endpoint (188). Using a developmental design this also investigated whether stereoscopic laparoscopic imaging impacted intraoperative TME performance. The feasibility of intraoperative QA at the centre of surgical intervention RCT study design was established. Again, variation in objective intraoperative performance was observed despite employing a number of credentialing and standardisation steps. The clinical implications of this finding were unclear and required investigation.

Equivalence between 2D and 3D arms was seen for all endpoints with no patient or surgeon benefits seen with 3D laparoscopic system use (188). These findings can inform the design of future 3D laparoscopic studies. Plane of mesorectal dissection is identified as an appropriate primary endpoint for a definitive 2D/3D laparoscopic TME RCT (188).

To investigate whether QA tools identify variability in surgical performance and if this impacts clinical outcomes, the two RCT datasets were combined and analysed. Objectively measured wide inter-surgeon performance and outcome variation was apparent and not explainable by demographic or other factors. When the credentialed surgeons were divided into skill quartiles based on their L-TMEpt scores, large and highly clinically relevant differences in nearly all perioperative and early outcome data were seen. Upper quartile surgeons obtained excellent laparoscopic TME specimens and half the morbidity of the other groups. Lower quartile surgeons

delivered inadequate results raising safety and ethical concerns. One of the main findings of this thesis is that the variation in results seen in complex surgical intervention studies are related to differences in intraoperative performance. It appears more appropriate to question who is performing laparoscopic TME and their demonstratable level of skill rather than the generalisability of the technique alone.

Finally, during the approximately 1800 hours of analysis performed in this research it was apparent that intraoperative QA could simpler and more clinically applicable. Further categorising the impact associated with observed error events would strengthen reporting. A novel system centred on the addition of an error severity category capturing the impact on the patient and/or procedure rather than simply count events was undertaken (230). This was developed with the assistance of a number of European MAS experts who defined their requirements and shaped the structure (230). The classification met acceptability and utility considerations as well as reliability and initial clinical validity assessments. The classification also records and allows analysis of intraoperative near misses which are typically overlooked (132, 166). Near misses are seen to be commonplace and closely associated with morbidity outcomes. Their identified frequency offers some advantages over capturing rarer but more serious events. External validity investigation is now required.

7.2 Discussion of hypothesis one: Quality assurance tools for laparoscopic total mesorectal excision can be developed and applied to research and clinical settings

In this thesis, previously reported MAS TME tools were reviewed with OCHRA and the developmental L-TMEpt taken forward (161). Their subsequent extensive application to specialist performed laparoscopic TME case video provided quality assurance through investigation of technique, intervention delivery and importantly to objectively measure surgical performance.

The reliability of the L-TMEpt tool is shown with levels of agreement comfortably exceeding those required for high-stakes and specialist assessment settings. Within the context of laparoscopic TME,

OCHRA had previously been shown to be valid and reliable (73, 159). These findings were confirmed here with the available large datasets. To further strengthen objective reporting, the new EAES intraoperative adverse event classification was developed to meet the identified need for a system applicable to the routine clinical practice setting (230). It is shown to be objective and reliable with initial supportive clinical validity evidence.

The literature search identified that intraoperative QA had been rarely performed in research or TME RCTs (161). To our knowledge the 176 laparoscopic TME cases successfully analysed with objective assessment tools presented here is one of the largest global experiences in any surgical procedure. Subsequently, the hypothesis can be supported as quality assurance tools for laparoscopic total mesorectal excision were developed and successfully applied.

7.3 Discussion of hypotheses two: Quality assurance tools can reliably assess operative technique and correlate with clinical outcomes following laparoscopic total mesorectal excision

In chapter 5, the combined analysis of the two trials confirmed that objectively assessed intraoperative QA performance measures were strongly associated with important histopathological and early clinical outcomes. These outcomes impacts upon the patient and their recovery and form established prognostic factors for oncological outcomes (41, 76, 77, 101). These findings hold potential implications for healthcare providers, surgical societies and regulatory bodies.

The credentialed specialist surgeons studied were previously assumed to represent a homogenous cohort (89, 188). Through quantification of their performance in 176 cases made possible with the application of intraoperative quality assurance, score variation was identified. When surgeon quartiles (based on their average scores) were analysed, outcome differences with significant clinical implications were observed between the groups. The large effect sizes meant very small numbers needed to treat were seen highlighting a new patient safety route with the potential to deliver

meaningful improvement. I am not aware of other rectal cancer quality improvement strategies that present effect sizes comparable to those reported here.

Therefore, the hypothesis that quality assurance tools can reliably assess operative technique and correlate with clinical outcomes following laparoscopic total mesorectal excision is accepted.

7.4 Discussion of specific objectives

The objectives of this thesis (section 1.16) are revisited with all considered to have been met.

Identify the reported tools for objective assessment of intraoperative laparoscopic TME performance

The PRISMA compliant systematic review described in chapter two successfully identified the reported assessment tools described for laparoscopic TME (161).

Apply objective assessment tools to randomised controlled trials to evaluate:

Tool reliability

The inter-rater, test retest and internal consistency of the L-TMEpt is demonstrated establishing the reliability of the tool for the assessment of specialist performed laparoscopic TME surgery. Excellent levels of agreement are reported and support ongoing use.

Concurrent validity

Comparison of OCHRA and L-TMEpt scores are shown to be concordant with moderate strength negative correlation reported. This is equal to the data from comparison of OCHRA with the colonic L-CAT tool performed within the LapCo programme (74).

Construct and clinical outcome validity

In chapter five, OCHRA error counts and L-TMEpt scores both correctly differentiated differing levels of performance. Although the quartiles were based upon L-TMEpt results, score difference were associated with OCHRA data as well as intraoperative, histopathological and 30-day patient morbidity outcome data. Large, clinically important effect sizes were seen between the quartiles.

High performance was associated with excellent outcomes significantly better than lower quartiles confirming construct validity.

7.5 Implications for the research setting and future research on this topic

The main findings of this thesis raise a number of avenues for ongoing study and there are a number of research implications for surgical trialists and methodologists. Surgical interventions are inherently variable and subject to influence from a number of factors. Through the systematic study of intraoperative performance, the surgeon is shown to represent a further major variable and one that can present major outcome implications. This is increasingly recognised in research and incorporated in frameworks such as the IDEAL collaboration (130, 139, 142, 143, 216, 217, 248). The findings here complement and support this effort as well as reinforce the need to identify and control for performance and its variation within surgical studies. This is of particular importance for interventional RCTs that provide high level evidence that influences guidelines and clinical practice.

The available methods used to assess the risk of bias in randomised trials were not designed for surgical intervention studies and their appropriateness has been questioned (144, 248). Heterogeneity statistics applied to meta-analyses are insufficiently explored in primary studies and individual level performance and outcome data (249). This thesis argues such research is required.

The identified variation in performance has implications for the design and delivery of future randomised trials. Credentialing based on prior case numbers, unit and surgeon volume or positions of esteem alone are shown to be inadequate and fails to eliminate heterogeneity. Pre-trial objective measurement of performance, ideally undertaken on multiple occasions, would be superior and provide data that could amend sample size calculations through incorporation of source of variance and adjustments for outcome clustering data. Performance level could be considered as a further stratification method when randomising patients in large RCTs. In cancer surgery, non-standard technique can result in a partial homogenisation of the trial arms (225). The Dutch D1 vs. D2 gastric cancer RCT is a high profile examples of this phenomenon where this issue could be considered to

invalidate the study (250). Increasing procedural and performance homogeneity arguably comes with a reduction in study generalisability and external validity to real world experience.

QA represents a systematic process investigating whether the service meets defined quality criteria. Intraoperative QA measures are shown to provide rich data to evaluate surgical intervention delivery, reasons for deviation and most importantly, insight into subsequent outcomes. It is unclear if uniform standardisation of technique and performance is achievable, but the data presented suggests its pursuit is justified and raising performance levels of the lower quartile or controversially even limiting their practice, could result in meaningful improvements for patients. As shown here, excellent or substandard performance will also alter the effect of the intervention and its outcomes but presently this is not considered in trial reporting (110, 139, 217, 225, 248).

The most pressing research need for this area regards predictive validity. Studies containing longitudinal designs and multiple assessments on individual surgeons undertaken by multiple assessors are urgently needed to explore intra-surgeon variation and its underlying origins as well as allow advanced reliability statistics to be applied. These are essential steps before the L-TMEpt could be considered for widespread specialist level assessment with important associated implications. Predictive validity is considered one of the hardest aspects of surgical education studies but represent a core consideration. The methodology used in this thesis prevented this analysis and its omission represents an important limitation of this work.

Now that measurable variation has been reported it is essential to investigate whether improvement is possible at the specialist level. The LapCo experience suggests that it is possible to deliver widespread improvement in experienced surgeons. However, this is not directly comparable as participants were learning laparoscopic colonic resection and cannot be considered proficient nor would meet the credentialing criteria applied within the two RCTs (71, 74, 131). Further educational research is indicated to explore whether the reasons behind poor outcomes relate to inappropriate

MDT decision making, surgical skill deficiency, human factors, proficiency learning curve considerations or combinations of these areas. An interactive qualitative approach is indicated.

It is not known if meaningful enhancement in specialist surgical performance is possible and whether like music, art, sport or thousands of other skills in life, some surgeons may simply be better than others. This would be a highly interesting study to undertake but as all surgeons are taught not born, I personally believe all have the capacity to improve. This follows the sporting analogy where participants from local parks up to the elite level all attempt to improve. Even when Olympians win a gold medal they continue to train and seek to progress. Targeted surgical training or coaching might represent routes towards improvement and could be individualised using OCHRA and L-TMEpt generated feedback. There is potential for such strategies to be incorporated into continuing professional development programmes for senior clinicians. This question is important as if meaningful alteration in performance is not possible, assessment may not be a worthwhile undertaking outside of research or dedicated summative training settings unless it were being used to restrict practice where insufficient performance is identified. This complex question requires focussed research.

The assessments described in this thesis were all performed in the research setting at a delayed timepoint remote from the operating theatre and responsible team. This provided homogeneity and a cleaner dataset but despite editorial calls for increasing use of video reviews it would not be suitable for wider future application (137, 138, 251, 252). The experimental nature of this research and the developmental nature of the L-TMEpt meant no feedback was given to individual surgeons. It is not known whether live assessment performed by a scrubbed trainer or assessor present in theatre would obtain comparable results as they may be subject to influence from the additional procedural information available and interaction and affiliation with the surgeon (229). The Hawthorne effect, where surgeon behaviour consciously or unconsciously changes due to the

awareness of being monitored, could theoretically result in differences from surgeon's usual performance. This has yet to be explored in surgical and MAS practice.

The importance for optimal pathological specimen quality has been highlighted throughout this thesis. Each RCT simplified the complex histopathological reports on each specimen to a number of binary endpoints to facilitate comparative analyses. If full reports where available, it would be possible to revisit the videos and correlate the pre and intraoperative events (or their omission) that led to inadequate results. Such research could be a powerful route to quality improvement.

This thesis relates to laparoscopic TME cases. In response to the acknowledged challenges it poses and the longstanding awareness that poor results are sometimes seen, a number of alternative procedures are available. Trans-anal TME seeks to avoid the difficulty of precise low pelvic surgery from abdominal port sites through performing these steps through an anal port and full thickness rectotomy approach to the distal mesorectal dissection (7, 78). A major RCT is currently recruiting to compare trans-anal and laparoscopic TME and is noteworthy that a number of intraoperative QA techniques including calls made in this thesis are present in the design (110, 248, 253). Robotic assisted surgery is also believed to present advantages in TME but this is not supported by high level evidence (81). To my knowledge, neither of these strategies has yet to explore intraoperative performance or quality assurance. Such work could be informative and help gauge the potential benefits and pitfalls of the techniques however the L-TMEpt and OCHRA methods and data presented here are specific to laparoscopic TME.

The EAES classification is ready for external validity assessment (230). Clinical validity is a further essential requirement that will determine uptake. It is hoped this report stimulates others to apply the classification to other procedures and specialities to explore its utility in widespread practice.

OCHRA provides a comprehensive analysis of the complex intraoperative period (73, 159, 169, 171, 172). The time and specialist surgical knowledge required prohibits its application beyond the research setting. I was excited to be approached by an expert machine learning and artificial

intelligence group with previous work in MAS assessment. Their proposal to explore automated OCHRA presents interesting possibilities. Such work may eventually lead to the ability to analyse all procedures and possibly even as they are performed (185). This could exponentially increase the use and nature of intraoperative quality assurance offering numerous quality improvement applications.

It would be appropriate and of real interest to perform qualitative research exploring patient (and stakeholder) viewpoints on the finding that surgical performance varies and holds implications. I am unsure how patients would react to the awareness and reporting of such data and how this might impact on their decision making including in providing informed consent and deciding on which surgeon or centre when seeking treatment. Clearly surgeons and representative bodies would need to be canvassed too but the potential conflicts of interest mean patient data should take priority.

7.6 Implications for laparoscopic TME clinical practice

Whilst it may be an obvious statement to say that good surgeons get good results, this thesis shows that the application of intraoperative QA objectively identifies good laparoscopic TME surgeons. It is demonstrated that there is a measurable difference in surgical performance of complex laparoscopic interventions, and this drives important outcome data. This conclusion raises a number of considerations within clinical practice and for associated regulatory bodies (124, 254).

As high performance is shown to associated with optimal results, it could be argued that its demonstration is in the best interests of patients and could be made mandatory before independent practice is permitted. This would be a considerably higher bar than current UK controls such as completion of training, professional examinations and guidelines (123, 133, 147). The ACPGBI clinical guidelines state that *"choice of rectal resection should be tailored to the individual patient, focussing on achieving R0 resection, low morbidity and restorative procedures in appropriate cases"* (147). Whilst this refers to the choice of operation, the findings of this thesis suggest that an amendment to include reference to the individual surgeon and their demonstratable level of performance might strengthen guidelines and their subsequent clinical impact.

In the latest UK national bowel cancer audit results 71% of major rectal cancer resections were started laparoscopically (9). This makes the findings of this thesis important to contemporary patients and surgeons. I believe the clinical validity data raises a number of challenging ethical questions. Although all MDTs and surgeons set out to deliver good care and obtain excellent outcomes, all the reported RCTs show that this was not achieved in a subset of patients (40). A number of surgeons may not possess sufficient skill to safely deliver laparoscopic TME as defined by their outcomes. Therefore, it could be argued that they should not be performing this surgery. This is an undoubtably tricky and thorny area but there is precedence as evidence based changes in practice such as the regionalisation of hepato-pancreatic-biliary surgical services following identification that surgeon and unit volumes were important factors (255-257). Subject to supportive data from further study as outlined, this thesis suggests such a review could be indicated in laparoscopic TME surgery and may be in the best interests of patients (258). This would require the participation of regulatory bodies and could be expected to raise controversy and resistance. The two tools presented here could assist these efforts and accredit performance.

Outside of the EAES classification design surgeon acceptability of intraoperative QA has not been sufficiently explored. Whilst objective assessment was supported by participants and assessors in the LapCo programme, they wished to be there and were being trained in new procedures that they wished to add to their practice (71, 74). Here, specialist independent practice is being assessed and one suggestion for objective assessment is more towards a policing role. Whilst offered with a patient focus and the best of intentions, the associated implications and consequences for those surgeons that do not display the required performance level and/or acceptable outcome data are significant. At the extreme end, restriction of practice could hold employment and financial implications with the possible loss of rectal cancer services from poorly performing centres. Clearly this work is a long way from this use, but the questions raised by these findings and the potential resulting consequences should not be ignored. After much consideration, I see no easy answers and it would appear appropriate to canvass the opinions of patients and support groups.

The 2D3D study showed that stereoscopic imaging gave no discernible impact on specialist intraoperative performance and associated outcome data (188). In contrast to previous generations of three dimensional systems no detrimental effects were seen suggesting equivalence, but the study does not support superiority over standard 2D platforms. Although health economics were not formally studied, the results are unlikely to justify the current higher purchase price nor provide supportive evidence for a business case.

It is unclear if the prior supportive evidence from trainee participants in the simulated or box trainer settings relates to a beneficial effect at that point of proficiency learning curves or the removal of bias due to the robust RCT methodology employed (197). It is noteworthy that a number of surgical technology trials objectively reported equivalence that ran against supportive opinion and results from lower quality studies (81, 218, 219). It remains my opinion that we have a duty to ensure all areas of our practice are evidence based and in the best interests of patients.

The EAES classification may help bridge the research presented here into routine clinical settings. The system is shown to be simple, reproducible and clinically applicable as requested by the expert participants (230). The proposed uses are potentially broadly generalisable and applicable across the breadth of MAS practice. It is hoped its presentation encourages other groups to study the classification within their procedures, practice and healthcare systems. The classification allows the study of near misses within procedures opening new research avenues akin to industrial practices.

7.7 Limitations of this thesis

In addition to the individual sections presented in each chapter, the results should be considered in the context of a number of themes that run through this thesis.

The results of the literature search defined all following steps of this research. Whilst a PRISMA compliant search and report was produced the string aimed to capture laparoscopic TME specific assessment methods. There are myriad described surgical assessment tools and some of the more

widely understood and applied methods may also have been applicable for use here but there were no reports on their use in laparoscopic TME surgery. However most, such as OSATS, were designed for training settings and therefore not suitable for application in this research (120, 122).

Many of the presented analyses were based upon the composition of the available video libraries. A significant proportion of ALaCaRT cases were either not recorded or not submitted to the central trial office representing protocol deviations. Although this did not affect the primary endpoint or other aspects of trial delivery, further cases would have strengthened this research and ensure there was no selection bias as theoretically surgeons may have not submitted poor performance cases. In personal communication from Mr Chris Brown (ALaCaRT lead statistician) this was not suspected and there was no difference in outcomes between video vs. no video laparoscopic cases. Based on the contracts agreed between the collaborating centres, no data on non-video cases was released to us preventing further exploration of this issue. Deviating from the CONSORT principles neither the ALaCaRT or 2D/3D studies reported any other data except number randomised meaning selection bias and non-consequential recruitment cannot be explored or fully excluded.

Whilst comparable pelvic analysis could be performed for both trials, the inclusion of all surgical phases would represent a more patient centred and appropriate strategy. Obtaining a beautiful TME dissection and cancer specimen would achieve high assessment scores and be valid for pathology endpoints. However, this might miss major issues in the abdominal tasks that would be likely to impact morbidity data. Despite this potential risk, strong clinical validity results were seen.

To our knowledge, this thesis contains the largest objective assessment of surgery to date. However, a number of important considerations remain. Entirely due to the extensive time required for multiple video analyses and the requirement for specialist surgical knowledge of laparoscopic TME surgery, outside of the stated inter-rater reliability assessments, analyses were performed by myself. This prevented the application of advanced reliability statistical assessments such as generalisability theory, exploratory factor analysis and source of variance studies. All of which demand multiple

assessors and ideally multiple assessments on the same individual at different timepoints. These results would be required before these methods could be applied to high-stake assessments.

This research was strengthened by the robust methods and data available from randomised trials. Unfortunately, no comment can be made on the predictive validity of the objective assessment scores obtained here. The 176 cases were performed by 48 surgeons limiting exploration of intra-surgeon variation. It cannot be assumed that surgeons display the same level between cases or what influences this as well as the distribution of an individual's scores. Predictive validity is another essential competent of summative or credentialing assessment as they must be proven to correlate with expected future performance (131).

These factors prevented the establishment of assessment score threshold or target associated with procedural competency and acceptable outcomes. As shown in figure 4.12, poor outcomes were seen across QA scores and not clustered as seen in LapCo (74). The associated detrimental impact on sensitivity and specificity meant thresholds required for clinical utility were not met.

Due to confidentiality considerations and the legal agreements between centres full histopathology reports from both RCTs were not available for review. Although the video analysis techniques applied are likely to capture events such as mesorectal fascial breech, combined analysis could have been informative on common reasons for inadequate specimens and its absence represents an important omission.

One goal of this research was to attempt to bring objective assessment out of the research setting towards routine practice. The L-TMEpt may achieve this but questions presently remain. Scores were defined through blinded video review. There is very little research exploring the comparison in scores between live in theatre scores and research settings (229). Assessors might be influenced by interaction or affinity for the surgeon and the extra awareness of relevant factors in each case. During such as a study It may be possible to simultaneously explore the reliability of L-TMEpt self-assessment as a further method towards routine application.

Video review primarily assesses the psychomotor domain and does not capture non-technical and human factors that can influence surgical performance (220, 227). These are very difficult to measure, control for or standardise between days, centres and surgeons. Distraction, fatigue, team composition and experience are examples of many factors that may influence surgeon performance and are likely to be present in all of the described trials as well as daily practice.

In view of the identified surgical variation and importance upon outcomes, it seems logical for trialists to attempt to define, measure and account for this in their studies. If competency assessment was mandatory for trial entry it could be expected to result in cleaner data that is more likely to truly assess the research question. However, the generalisability of any findings could be expected to decrease as surgeons that take up the technique would be unlikely to achieve comparable results. This raises ethical issues and the need for safe introduction of novel techniques and technologies. Lessons have been learnt from prior rapid role out of new procedures notably the initial high bile duct injury rate when laparoscopic cholecystectomy was taken up (57). Successes include the LapCo experience and structured programmes are underway for novel techniques such as trans-anal TME (160).

The EAES intraoperative error classification is still at an early stage and may not suitable for widespread clinical use yet. Similar to objective assessment as described above, a large number of assessors are required to establish reliability especially if the future use of the classification depends on self-reporting. Acceptability in view of potential medicolegal issues from self-reporting errors is needed from a variety of healthcare settings as well as legal advice from multiple jurisdictions are required to reassure surgeons. The classification is intended as a way to assist reporting, patient care and quality improvement not to provide evidence of inadequate performance. It is likely that educational initiatives would be required to disseminate this important fact.

7.8 Personal reflections

This work was a major undertaking and I believe I will take much forward from its successes but probably more from the challenges faced. It is hoped this text reflects the breadth, depth and quality of the academic and clinical training and opportunities this experience has provided. I now know that I can, have and will again design and deliver surgical research studies. It was a privilege to be at the centre of every step of the 2D3D trial from the brainstorming moment that "we need to do a multicentre trial" through the subsequent design, funding, approvals, delivery and reporting phases.

I have long given up estimating the number of hours of video review performed here with no doubt it is comfortably above 1800. However, I recognise that my operative training has been massively enhanced though repeated structured scrutiny of expert performance. Stamina is an underrated quality and it was no small feat just to capture these cases let alone analyse them, then again and then once more. In my opinion the detail obtained from structured video review is second to none and I hope efforts in artificial intelligence and machine learning are fruitful as future objective analysis of all our interventions could represent a huge driver for quality improvement (185).

As I return to clinical training, I am acutely aware that my target is not to tick off checklists and meet form and case number requirements for deanery and completion of training needs. It must be to aspire for excellence as this is shown to directly benefit those afflicted by colorectal cancer. Achieving the consultant level is only a step on the journey not the destination. This thesis shows that consultant surgeons deliver differing performance levels and this matters for patients.

7.9 Concluding remarks

The development and application of intraoperative quality assurance to two laparoscopic TME randomised controlled trials of credentialed surgeons allowed performance to be objectively measured and assess whether procedures meet quality criteria. Variation in technical skill is identified and shown to strongly link to clinical and pathological outcomes. This research confirms

that good surgeons deliver high performance with resulting good outcomes following laparoscopic TME. The intraoperative period should form the focus of complex surgical intervention studies and quality improvement initiatives.

Reference list

1. Salerno G, Sinnatamby C, Branagan G, Daniels IR, Heald RJ, Moran BJ. Defining the rectum: surgically, radiologically and anatomically. Colorectal Dis. 2006;8 Suppl 3:5-9.
2. McMullen TP, Easson AM, Cohen Z, Swallow CJ. The investigation of primary rectal cancer by surgeons: current pattern of practice. Can J Surg. 2005;48(1):19-26.
3. Wasserman MA, McGee MF, Helenowski IB, Halverson AL, Boller AM, Stryker SJ. The anthropometric definition of the rectum is highly variable. Int J Colorectal Dis. 2016;31(2):189-95.
4. Kenig J, Richter P. Definition of the rectum and level of the peritoneal reflection - still a matter of debate? Wideochir Inne Tech Maloinwazyjne. 2013;8(3):183-6.
5. Heald RJ, Moran BJ. Embryology and anatomy of the rectum. Semin Surg Oncol. 1998;15(2):66-71.
6. Lee JM, Kim NK. Essential Anatomy of the Anorectum for Colorectal Surgeons Focused on the Gross Anatomy and Histologic Findings. Ann Coloproctol. 2018;34(2):59-71.
7. Penna M, Hompes R, Arnold S, Wynn G, Austin R, Warusavitarne J, et al. Transanal Total Mesorectal Excision: International Registry Results of the First 720 Cases. Ann Surg. 2016.
8. Wedel T. Anatomy of the Colon, Rectum, Anus, and Pelvic Floor. In: Herold A, Lehur P-A, Matzel KE, O'Connell PR, editors. Coloproctology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2017. p. 7-22.
9. (HQIP) THQIP. National Bowel Cancer Audit Annual Report 20172017 6th July 2018.
10. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. Br J Surg. 2002;89(7):845-60.
11. Weitz J, Koch M, Debus J, Hohler T, Galle PR, Buchler MW. Colorectal cancer. Lancet. 2005;365(9454):153-65.
12. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350-6.
13. Dukes CE. The Surgical Pathology of Rectal Cancer. J Clin Pathol. 1949;2(2):95-8.
14. Loughrey M BQ, P; Shephard, N.A. Standards and datasets for reporting cancers - Dataset for histopathological reporting of colorectal cancer. London: Royal College of Pathologists; 2017 December 2017. Contract No.: G049.
15. NICE. Colorectal cancer: diagnosis and Colorectal cancer: diagnosis and management (CG131). National Institute for Health and Care Excellence; 2011 1st November 2011.
16. Doyle VJ, Bateman AC. Colorectal cancer staging using TNM 7: is it time to use this new staging system? J Clin Pathol. 2012;65(4):372-4.
17. Ueno H, Mochizuki H, Akagi Y, Kusumi T, Yamada K, Ikegami M, et al. Optimal colorectal cancer staging criteria in TNM classification. J Clin Oncol. 2012;30(13):1519-26.
18. MacFarlane JK, Ryall RD, Heald RJ. Mesorectal excision for rectal cancer. Lancet. 1993;341(8843):457-60.
19. Adam IJ, Mohamdee MO, Martin IG, Scott N, Finan PJ, Johnston D, et al. Role of circumferential margin involvement in the local recurrence of rectal cancer. Lancet. 1994;344(8924):707-11.
20. Quirke P, Durdey P, Dixon MF, Williams NS. Local recurrence of rectal adenocarcinoma due to inadequate surgical resection. Histopathological study of lateral tumour spread and surgical excision. Lancet. 1986;2(8514):996-9.

21. Quirke P. Training and quality assurance for rectal cancer: 20 years of data is enough. Lancet Oncol. 2003;4(11):695-702.
22. Harji DP, Koh C, Solomon M, Velikova G, Sagar PM, Brown J. Development of a conceptual framework of health-related quality of life in locally recurrent rectal cancer. Colorectal Dis. 2015;17(11):954-64.
23. Harji DP, Griffiths B, Velikova G, Sagar PM, Brown J. Systematic review of health-related quality of life issues in locally recurrent rectal cancer. J Surg Oncol. 2015;111(4):431-8.
24. Platell C, Spilsbury K. Influence of local recurrence on survival in patients with rectal cancer. ANZ J Surg. 2014;84(1-2):85-90.
25. van den Brink M, Stiggelbout AM, van den Hout WB, Kievit J, Klein Kranenbarg E, Marijnen CA, et al. Clinical nature and prognosis of locally recurrent rectal cancer after total mesorectal excision with or without preoperative radiotherapy. J Clin Oncol. 2004;22(19):3958-64.
26. Group MS. Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study. BMJ. 2006;333(7572):779.
27. Battersby NJ, How P, Moran B, Stelzner S, West NP, Branagan G, et al. Prospective Validation of a Low Rectal Cancer Magnetic Resonance Imaging Staging System and Development of a Local Recurrence Risk Stratification Model: The MERCURY II Study. Ann Surg. 2016;263(4):751-60.
28. Al-Sukhni E, Milot L, Fruitman M, Beyene J, Victor JC, Schmocker S, et al. Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis. Ann Surg Oncol. 2012;19(7):2212-23.
29. Taylor FG, Quirke P, Heald RJ, Moran BJ, Blomqvist L, Swift IR, et al. Preoperative magnetic resonance imaging assessment of circumferential resection margin predicts disease-free survival and local recurrence: 5-year follow-up results of the MERCURY study. J Clin Oncol. 2014;32(1):34-43.
30. NICE. Colorectal cancer - Quality Standard (QS20). National Institute for Health and Care Excellence; 2012 13 August 2012.
31. Cunningham C, Leong K, Clark S, Plumb A, Taylor S, Geh I, et al. Association of Coloproctology of Great Britain & Ireland (ACPGBI): Guidelines for the Management of Cancer of the Colon, Rectum and Anus (2017) - Diagnosis, Investigations and Screening. Colorectal Dis. 2017;19 Suppl 1:9-17.
32. Sauer R, Becker H, Hohenberger W, Rodel C, Wittekind C, Fietkau R, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med. 2004;351(17):1731-40.
33. Sebag-Montefiore D, Stephens RJ, Steele R, Monson J, Grieve R, Khanna S, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): a multicentre, randomised trial. Lancet. 2009;373(9666):811-20.
34. De Caluwe L, Van Nieuwenhove Y, Ceelen WP. Preoperative chemoradiation versus radiation alone for stage II and III resectable rectal cancer. Cochrane Database Syst Rev. 2013(2):CD006041.
35. Wolthuis AM, Penninckx F, Haustermans K, De Hertogh G, Fieuws S, Van Cutsem E, et al. Impact of interval between neoadjuvant chemoradiotherapy and TME for locally advanced rectal cancer on pathologic response and oncologic outcome. Ann Surg Oncol. 2012;19(9):2833-41.
36. Heald RJ, Husband EM, Ryall RD. The mesorectum in rectal cancer surgery--the clue to pelvic recurrence? Br J Surg. 1982;69(10):613-6.
37. Heald RJ, Ryall RD. Recurrence and survival after total mesorectal excision for rectal cancer. Lancet. 1986;1(8496):1479-82.

38. Miskovic D, Foster J, Agha A, Delaney CP, Francis N, Hasegawa H, et al. Standardization of laparoscopic total mesorectal excision for rectal cancer: a structured international expert consensus. Ann Surg. 2015;261(4):716-22.
39. Fleshman J, Branda ME, Sargent DJ, Boller AM, George VV, Abbas MA, et al. Disease-free Survival and Local Recurrence for Laparoscopic Resection Compared With Open Resection of Stage II to III Rectal Cancer: Follow-up Results of the ACOSOG Z6051 Randomized Controlled Trial. Ann Surg. 2018.
40. Vennix S, Pelzers L, Bouvy N, Beets GL, Pierie JP, Wiggers T, et al. Laparoscopic versus open total mesorectal excision for rectal cancer. Cochrane Database Syst Rev. 2014(4):CD005200.
41. Kitz J, Fokas E, Beissbarth T, Strobel P, Wittekind C, Hartmann A, et al. Association of Plane of Total Mesorectal Excision With Prognosis of Rectal Cancer: Secondary Analysis of the CAO/ARO/AIO-04 Phase 3 Randomized Clinical Trial. JAMA Surg. 2018:e181607.
42. Kapiteijn E, Putter H, van de Velde CJ, Cooperative investigators of the Dutch ColoRectal Cancer G. Impact of the introduction and training of total mesorectal excision on recurrence and survival in rectal cancer in The Netherlands. Br J Surg. 2002;89(9):1142-9.
43. Martling A, Holm T, Rutqvist LE, Johansson H, Moran BJ, Heald RJ, et al. Impact of a surgical training programme on rectal cancer outcomes in Stockholm. Br J Surg. 2005;92(2):225-9.
44. Battersby NJ, Juul T, Christensen P, Janjua AZ, Branagan G, Emmertsen KJ, et al. Predicting the Risk of Bowel-Related Quality-of-Life Impairment After Restorative Resection for Rectal Cancer: A Multicenter Cross-Sectional Study. Dis Colon Rectum. 2016;59(4):270-80.
45. Trenti L, Galvez A, Biondo S, Solis A, Vallribera-Valls F, Espin-Basany E, et al. Quality of life and anterior resection syndrome after surgery for mid to low rectal cancer: A cross-sectional study. Eur J Surg Oncol. 2018;44(7):1031-9.
46. Chapman SJ, Bolton WS, Corrigan N, Young N, Jayne DG. A Cross-Sectional Review of Reporting Variation in Postoperative Bowel Dysfunction After Rectal Cancer Surgery. Dis Colon Rectum. 2017;60(2):240-7.
47. Celentano V, Cohen R, Warusavitarne J, Faiz O, Chand M. Sexual dysfunction following rectal cancer surgery. Int J Colorectal Dis. 2017;32(11):1523-30.
48. Church JM, Raudkivi PJ, Hill GL. The surgical anatomy of the rectum--a review with particular relevance to the hazards of rectal mobilisation. Int J Colorectal Dis. 1987;2(3):158-66.
49. Lindsey I, Guy RJ, Warren BF, Mortensen NJ. Anatomy of Denonvilliers' fascia and pelvic nerves, impotence, and implications for the colorectal surgeon. Br J Surg. 2000;87(10):1288-99.
50. Runkel N, Reiser H. Nerve-oriented mesorectal excision (NOME): autonomic nerves as landmarks for laparoscopic rectal resection. Int J Colorectal Dis. 2013;28(10):1367-75.
51. Schwab K, Smith R, Brown V, Whyte M, Jourdan I. Evolution of stereoscopic imaging in surgery and recent advances. World J Gastrointest Endosc. 2017;9(8):368-77.
52. Schwenk W, Haase O, Neudecker J, Muller JM. Short term benefits for laparoscopic colorectal resection. Cochrane Database Syst Rev. 2005(3):CD003145.
53. Veldkamp R, Gholghesaei M, Bonjer HJ, Meijer DW, Buunen M, Jeekel J, et al. Laparoscopic resection of colon Cancer: consensus of the European Association of Endoscopic Surgery (EAES). Surg Endosc. 2004;18(8):1163-85.
54. Veenhof AA, Vlug MS, van der Pas MH, Sietses C, van der Peet DL, de Lange-de Klerk ES, et al. Surgical stress response and postoperative immune function after laparoscopy or open surgery with fast track or standard perioperative care: a randomized trial. Ann Surg. 2012;255(2):216-21.

55. Beal GM, Rogers EM, Bohlen JM. Validity of the concept of stages in the adoption process. Rural Sociology. 1957;22(2):166-8.
56. Miskovic D, Ni M, Wyles SM, Tekkis P, Hanna GB. Learning curve and case selection in laparoscopic colorectal surgery: systematic review and international multicenter analysis of 4852 cases. Dis Colon Rectum. 2012;55(12):1300-10.
57. MacFadyen BV, Jr., Vecchio R, Ricardo AE, Mathis CR. Bile duct injury after laparoscopic cholecystectomy. The United States experience. Surg Endosc. 1998;12(4):315-21.
58. CholeS Study Group WMRC. Population-based cohort study of outcomes following cholecystectomy for benign gallbladder diseases. Br J Surg. 2016;103(12):1704-15.
59. Noble F, Kelly JJ, Bailey IS, Byrne JP, Underwood TJ, South Coast Cancer Collaboration O-G. A prospective comparison of totally minimally invasive versus open Ivor Lewis esophagectomy. Dis Esophagus. 2013;26(3):263-71.
60. Wang C, Sun P, Li J, Yang W, Yang J, Feng Z, et al. Strategies of laparoscopic thyroidectomy for treatment of substernal goiter via areola approach. Surg Endosc. 2016;30(11):4721-30.
61. Robertson L, Nandhra S. Laparoscopic surgery for elective abdominal aortic aneurysm repair. Cochrane Database Syst Rev. 2017;5:CD012302.
62. Jacobs M, Verdeja JC, Goldstein HS. Minimally invasive colon resection (laparoscopic colectomy). Surg Laparosc Endosc. 1991;1(3):144-50.
63. Wexner SD, Cohen SM. Port site metastases after laparoscopic colorectal surgery for cure of malignancy. Br J Surg. 1995;82(3):295-8.
64. Chapman AE, Levitt MD, Hewett P, Woods R, Sheiner H, Maddern GJ. Laparoscopic-assisted resection of colorectal malignancies: a systematic review. Ann Surg. 2001;234(5):590-606.
65. Kuhry E, Schwenk WF, Gaupset R, Romild U, Bonjer HJ. Long-term results of laparoscopic colorectal cancer resection. Cochrane Database Syst Rev. 2008(2):CD003432.
66. A comparison of laparoscopically assisted and open colectomy for colon cancer. N Engl J Med. 2004;350(20):2050-9.
67. Veldkamp R, Kuhry E, Hop WC, Jeekel J, Kazemier G, Bonjer HJ, et al. Laparoscopic surgery versus open surgery for colon cancer: short-term outcomes of a randomised trial. Lancet Oncol. 2005;6(7):477-84.
68. Bonjer HJ, Hop WC, Nelson H, Sargent DJ, Lacy AM, Castells A, et al. Laparoscopically assisted vs open colectomy for colon cancer: a meta-analysis. Arch Surg. 2007;142(3):298-303.
69. Buunen M, Veldkamp R, Hop WC, Kuhry E, Jeekel J, Haglind E, et al. Survival after laparoscopic surgery versus open surgery for colon cancer: long-term outcome of a randomised clinical trial. Lancet Oncol. 2009;10(1):44-52.
70. NICE. Laparoscopic surgery for colorectal Laparoscopic surgery for colorectal cancer (TA105). National Institute for Health and Care Excellence; 2006 23 August 2006. Report No.: TA105.
71. Coleman MG, Hanna GB, Kennedy R. The National Training Programme for Laparoscopic Colorectal Surgery in England: a new training paradigm. Colorectal Dis. 2011;13(6):614-6.
72. Miskovic D, Wyles SM, Carter F, Coleman MG, Hanna GB. Development, validation and implementation of a monitoring tool for training in laparoscopic colorectal surgery in the English National Training Program. Surg Endosc. 2011;25(4):1136-42.
73. Miskovic D, Ni M, Wyles SM, Parvaiz A, Hanna GB. Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. Surg Endosc. 2012;26(3):796-803.

74. Miskovic D, Ni M, Wyles SM, Kennedy RH, Francis NK, Parvaiz A, et al. Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England. Ann Surg. 2013;257(3):476-82.
75. Leroy J, Jamali F, Forbes L, Smith M, Rubino F, Mutter D, et al. Laparoscopic total mesorectal excision (TME) for rectal cancer surgery: long-term outcomes. Surg Endosc. 2004;18(2):281-9.
76. Nagtegaal ID, van de Velde CJ, van der Worp E, Kapiteijn E, Quirke P, van Krieken JH. Macroscopic evaluation of rectal cancer resection specimen: clinical significance of the pathologist in quality control. J Clin Oncol. 2002;20(7):1729-34.
77. Quirke P, Steele R, Monson J, Grieve R, Khanna S, Couture J, et al. Effect of the plane of surgery achieved on local recurrence in patients with operable rectal cancer: a prospective study using data from the MRC CR07 and NCIC-CTG CO16 randomised clinical trial. Lancet. 2009;373(9666):821-8.
78. Simillis C, Hompes R, Penna M, Rasheed S, Tekkis PP. A systematic review of transanal total mesorectal excision: is this the future of rectal cancer surgery? Colorectal Dis. 2016;18(1):19-36.
79. Borstlap WA, Tanis PJ, Koedam TW, Marijnen CA, Cunningham C, Dekker E, et al. A multi-centred randomised trial of radical surgery versus adjuvant chemoradiotherapy after local excision for early rectal cancer. BMC Cancer. 2016;16:513.
80. Rombouts AJM, Al-Najami I, Abbott NL, Appelt A, Baatrup G, Bach S, et al. Can we Save the rectum by watchful waiting or TransAnal microsurgery following (chemo) Radiotherapy versus Total mesorectal excision for early REctal Cancer (STAR-TREC study)?: protocol for a multicentre, randomised feasibility study. BMJ Open. 2017;7(12):e019474.
81. Jayne D, Pigazzi A, Marshall H, Croft J, Corrigan N, Copeland J, et al. Effect of Robotic-Assisted vs Conventional Laparoscopic Surgery on Risk of Conversion to Open Laparotomy Among Patients Undergoing Resection for Rectal Cancer: The ROLARR Randomized Clinical Trial. JAMA. 2017;318(16):1569-80.
82. Guillou PJ, Quirke P, Thorpe H, Walker J, Jayne DG, Smith AM, et al. Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer (MRC CLASICC trial): multicentre, randomised controlled trial. Lancet. 2005;365(9472):1718-26.
83. Jayne DG, Guillou PJ, Thorpe H, Quirke P, Copeland J, Smith AM, et al. Randomized trial of laparoscopic-assisted resection of colorectal carcinoma: 3-year results of the UK MRC CLASICC Trial Group. J Clin Oncol. 2007;25(21):3061-8.
84. van der Pas MH, Haglind E, Cuesta MA, Furst A, Lacy AM, Hop WC, et al. Laparoscopic versus open surgery for rectal cancer (COLOR II): short-term outcomes of a randomised, phase 3 trial. Lancet Oncol. 2013;14(3):210-8.
85. Bonjer HJ, Deijen CL, Abis GA, Cuesta MA, van der Pas MH, ES dL-dK, et al. A randomized trial of laparoscopic versus open surgery for rectal cancer. N Engl J Med. 2015;372(14):1324-32.
86. Kang SB, Park JW, Jeong SY, Nam BH, Choi HS, Kim DW, et al. Open versus laparoscopic surgery for mid or low rectal cancer after neoadjuvant chemoradiotherapy (COREAN trial): short-term outcomes of an open-label randomised controlled trial. Lancet Oncol. 2010;11(7):637-45.
87. Jeong SY, Park JW, Nam BH, Kim S, Kang SB, Lim SB, et al. Open versus laparoscopic surgery for mid-rectal or low-rectal cancer after neoadjuvant chemoradiotherapy (COREAN trial): survival outcomes of an open-label, non-inferiority, randomised controlled trial. Lancet Oncol. 2014;15(7):767-74.

88. Fleshman J, Branda M, Sargent DJ, Boller AM, George V, Abbas M, et al. Effect of Laparoscopic-Assisted Resection vs Open Resection of Stage II or III Rectal Cancer on Pathologic Outcomes: The ACOSOG Z6051 Randomized Clinical Trial. JAMA. 2015;314(13):1346-55.
89. Stevenson AR, Solomon MJ, Lumley JW, Hewett P, Clouston AD, Gebski VJ, et al. Effect of Laparoscopic-Assisted Resection vs Open Resection on Pathological Outcomes in Rectal Cancer: The ALaCaRT Randomized Clinical Trial. JAMA. 2015;314(13):1356-63.
90. Stevenson ARL, Solomon MJ, Brown CSB, Lumley JW, Hewett P, Clouston AD, et al. Disease-free Survival and Local Recurrence After Laparoscopic-assisted Resection or Open Resection for Rectal Cancer: The Australasian Laparoscopic Cancer of the Rectum Randomized Clinical Trial. Ann Surg. 2018.
91. Martinez-Perez A, Carra MC, Brunetti F, de'Angelis N. Pathologic Outcomes of Laparoscopic vs Open Mesorectal Excision for Rectal Cancer: A Systematic Review and Meta-analysis. JAMA Surg. 2017;152(4):e165665.
92. Martinez-Perez A, Carra MC, Brunetti F, de'Angelis N. Short-term clinical outcomes of laparoscopic vs open rectal excision for rectal cancer: A systematic review and meta-analysis. World J Gastroenterol. 2017;23(44):7906-16.
93. Acuna SA, Chesney TR, Ramjist JK, Shah PS, Kennedy ED, Baxter NN. Laparoscopic Versus Open Resection for Rectal Cancer: A Noninferiority Meta-analysis of Quality of Surgical Resection Outcomes. Ann Surg. 2018.
94. Zhou ZG, Hu M, Li Y, Lei WZ, Yu YY, Cheng Z, et al. Laparoscopic versus open total mesorectal excision with anal sphincter preservation for low rectal cancer. Surg Endosc. 2004;18(8):1211-5.
95. Braga M, Frasson M, Vignali A, Zuliani W, Capretti G, Di C, V. Laparoscopic resection in rectal cancer patients: outcome and cost-benefit analysis. Dis Colon Rectum. 2007;50(4):464-71.
96. Ng SS, Leung KL, Lee JF, Yiu RY, Li JC, Teoh AY, et al. Laparoscopic-assisted versus open abdominoperineal resection for low rectal cancer: a prospective randomized trial. Ann Surg Oncol. 2008;15(9):2418-25.
97. Lujan J, Valero G, Hernandez Q, Sanchez A, Frutos MD, Parrilla P. Randomized clinical trial comparing laparoscopic and open surgery in patients with rectal cancer. Br J Surg. 2009;96(9):982-9.
98. Liang X, Hou S, Liu H, Li Y, Jiang B, Bai W, et al. Effectiveness and safety of laparoscopic resection versus open surgery in patients with rectal cancer: a randomized, controlled trial from China. J Laparoendosc Adv Surg Tech A. 2011;21(5):381-5.
99. Gong J, Shi DB, Li XX, Cai SJ, Guan ZQ, Xu Y. Short-term outcomes of laparoscopic total mesorectal excision compared to open surgery. World J Gastroenterol. 2012;18(48):7308-13.
100. Ng SS, Lee JF, Yiu RY, Li JC, Hon SS, Mak TW, et al. Laparoscopic-assisted versus open total mesorectal excision with anal sphincter preservation for mid and low rectal cancer: a prospective, randomized trial. Surg Endosc. 2014;28(1):297-306.
101. Leonard D, Penninckx F, Laenen A, Kartheuser A. Scoring the quality of total mesorectal excision for the prediction of cancer-specific outcome. Colorectal Dis. 2015;17(5):O115-O22.
102. Rickles AS, Dietz DW, Chang GJ, Wexner SD, Berho ME, Remzi FH, et al. High Rate of Positive Circumferential Resection Margins Following Rectal Cancer Surgery: A Call to Action. Ann Surg. 2015;262(6):891-8.
103. Moran BJ, Holm T, Brannagan G, Chave H, Quirke P, West N, et al. The English national low rectal cancer development programme: key messages and future perspectives. Colorectal Dis. 2014;16(3):173-8.

104. Foster JD, Gash KJ, Carter FJ, West NP, Acheson AG, Horgan AF, et al. Development and evaluation of a cadaveric training curriculum for low rectal cancer surgery in the English LOREC National Development Programme. Colorectal Dis. 2014;16(9):O308-19.
105. Bloom BSH, J.T; Mandaus, G.F. Handbook on formative and summative evaluation of student learning: McGraw Hill; 1971.
106. Cuschieri GBHNKFA. Psychomotor ability testing and human reliability analysis (HRA) in surgical practice. Minimally Invasive Therapy & Allied Technologies. 2001;10(3):181-95.
107. Reznick RK. Teaching and testing technical skills. Am J Surg. 1993;165(3):358-61.
108. Cohen LM, L; Morrison, K. Research methods in education. 7th ed: Routledge; 2011.
109. Mackenzie H, Miskovic D, Ni M, Parvaiz A, Acheson AG, Jenkins JT, et al. Clinical and educational proficiency gain of supervised laparoscopic colorectal surgical trainees. Surg Endosc. 2013;27(8):2704-11.
110. Foster JD, Mackenzie H, Nelson H, Hanna GB, Francis NK. Methods of quality assurance in multicenter trials in laparoscopic colorectal surgery: a systematic review. Ann Surg. 2014;260(2):220-9.
111. Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. BMJ. 1999;318(7188):887-8.
112. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. Am J Surg. 2011;202(4):469-80 e6.
113. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. BMJ. 2003;327(7422):1032-7.
114. Cuschieri A, Francis N, Crosby J, Hanna GB. What do master surgeons think of surgical competence and revalidation? Am J Surg. 2001;182(2):110-6.
115. Black P, Wiliam D. Developing the theory of formative assessment. Educational Assessment, evaluation and accountability. 2009;21(1):5-31.
116. Jenkins JT, Currie A, Sala S, Kennedy RH. A multi-modal approach to training in laparoscopic colorectal surgery accelerates proficiency gain. Surg Endosc. 2016;30(7):3007-13.
117. ISCP. Procedure Based Assessment https://www.iscp.ac.uk/surgical/assessment_pba.aspx 2016 [updated 2016.
118. Beard J, Rowley D, Bussey M, Pitts D. Workplace-based assessment: assessing technical skill throughout the continuum of surgical training. ANZ J Surg. 2009;79(3):148-53.
119. Szasz P, Louridas M, Harris KA, Aggarwal R, Grantcharov TP. Assessing Technical Competence in Surgical Trainees: A Systematic Review. Ann Surg. 2015;261(6):1046-55.
120. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. Br J Surg. 2010;97(7):972-87.
121. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg. 2005;190(1):107-13.
122. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273-8.
123. Beard JD, Marriott J, Purdie H, Crossley J. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. Health Technol Assess. 2011;15(1):i-xxi, 1-162.
124. Clinical outcomes publication 2017 [Internet]. 2017 [cited 05/09/2018]. Available from: www.acpgbi.org.uk/surgeon-outcomes/.
125. Outcomes [Internet]. 2015 [cited 06/08/2018]. Available from: www.scts.org/outcomes/.

126. Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA. Ceiling effect in technical skills of surgical residents. Am J Surg. 2004;188(3):294-300.
127. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9 Suppl):S63-7.
128. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc. 2003;17(10):1525-9.
129. Foster JD. Objective Assessment of laparoscopic rectal cancer surgery. London: Imperial College London; 2015.
130. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med. 2013;369(15):1434-42.
131. Mackenzie H, Ni M, Miskovic D, Motson RW, Gudgeon M, Khan Z, et al. Clinical validity of consultant technical skills assessment in the English National Training Programme for Laparoscopic Colorectal Surgery. Br J Surg. 2015;102(8):991-7.
132. Bonrath EM, Gordon LE, Grantcharov TP. Characterising 'near miss' events in complex laparoscopic surgery through video analysis. BMJ Qual Saf. 2015;24(8):516-21.
133. Marriott J, Purdie H, Crossley J, Beard JD. Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. Br J Surg. 2011;98(3):450-7.
134. Beard J. Workplace-based assessment: the need for continued evaluation and refinement. Surgeon. 2011;9 Suppl 1:S12-3.
135. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med. 1998;73(9):993-7.
136. ISCP. Intercollegiate Surgical Curriculum Project [Available from: www.iscp.ac.uk.
137. Langerman A, Grantcharov TP. Are We Ready for Our Close-up?: Why and How We Must Embrace Video in the OR. Ann Surg. 2017;266(6):934-6.
138. Dimick JB, Varban OA. Surgical video analysis: an emerging tool for improving surgeon performance. BMJ Qual Saf. 2015;24(8):490-1.
139. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374(9695):1105-12.
140. Foster JD, Ewings P, Falk S, Cooper EJ, Roach H, West NP, et al. Surgical timing after chemoradiotherapy for rectal cancer, analysis of technique (STARRCAT): results of a feasibility multi-centre randomized controlled trial. Tech Coloproctol. 2016.
141. Dixon AF, C; Harrison, T. Preparing for the Francis report - How to quality assure in the NHS. The King's Fund; 2012 July 2012.
142. Fecso AB, Bhatti JA, Stotland PK, Quereshy FA, Grantcharov TP. Technical Performance as a Predictor of Clinical Outcomes in Laparoscopic Gastric Cancer Surgery. Ann Surg. 2018.
143. Fecso AB, Szasz P, Kerezov G, Grantcharov TP. The Effect of Technical Performance on Patient Outcomes in Surgery: A Systematic Review. Ann Surg. 2017;265(3):492-501.
144. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, Group CN. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. Ann Intern Med. 2017;167(1):40-7.
145. D'Souza N, Balyasnikova S, Tudkya V, Lord A, Shaw A, Abulafi M, et al. Variation in Landmarks for the Rectum: An MRI Study. Colorectal Dis. 2018.

146. Monson JR, Weiser MR, Buie WD, Chang GJ, Rafferty JF, Buie WD, et al. Practice parameters for the management of rectal cancer (revised). Dis Colon Rectum. 2013;56(5):535-50.
147. Moran B, Cunningham C, Singh T, Sagar P, Bradbury J, Geh I, et al. Association of Coloproctology of Great Britain & Ireland (ACPGBI): Guidelines for the Management of Cancer of the Colon, Rectum and Anus (2017) - Surgical Management. Colorectal Dis. 2017;19 Suppl 1:18-36.
148. Faiz O. Association of Coloproctology of Great Britain & Ireland (ACPGBI): Guidelines for the Management of Cancer of the Colon, Rectum and Anus (2017) - Audit and Outcome Reporting. Colorectal Dis. 2017;19 Suppl 1:71-3.
149. Keikes L, Koopman M, Tanis PJ, Lemmens V, Punt CJA, van Oijen MGH. Evaluating the scientific basis of quality indicators in colorectal cancer care: A systematic review. Eur J Cancer. 2017;86:166-77.
150. Stulberg JJ, Delaney CP, Neuhauser DV, Aron DC, Fu P, Koroukian SM. Adherence to surgical care improvement project measures and the association with postoperative infections. JAMA. 2010;303(24):2479-85.
151. Fleshman J, Sargent DJ, Green E, Anvari M, Stryker SJ, Beart RW, Jr., et al. Laparoscopic colectomy for cancer is not inferior to open surgery based on 5-year data from the COST Study Group trial. Ann Surg. 2007;246(4):655-62; discussion 62-4.
152. Miskovic D. Proficiency gain and competency assessment in laparoscopic colorectal surgery: Imperial College London; 2012.
153. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
154. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52(6):377-84.
155. JPT H, DG A, JAC S. Assessing risk of bias in included studies. In: JPT H, S G, editors. Cochrane Handbook for Systematic Reviews of Interventions. 5.1.0 ed: The Cochrane Collaboration; 2011.
156. Sarker SK, Kumar I, Delaney C. Assessing operative performance in advanced laparoscopic colorectal surgery. World J Surg. 2010;34(7):1594-603.
157. Sarker SK, Delaney C. Feasibility of self-appraisal in assessing operative performance in advanced laparoscopic colorectal surgery. Colorectal Dis. 2011;13(7):805-10.
158. Costanzi A, Rigamonti L, Mari GM, Miranda A, Crippa J, Berardi V, et al. A prospective video-controlled study of genito-urinary disorders in 35 consecutive laparoscopic TMEs for rectal cancer. Surg Endosc. 2015;29(7):1721-8.
159. Foster JD, Miskovic D, Allison AS, Conti JA, Ockrim J, Cooper EJ, et al. Application of objective clinical human reliability analysis (OCHRA) in assessment of technical performance in laparoscopic rectal cancer surgery. Tech Coloproctol. 2016;20(6):361-7.
160. Penna M, Hompes R, Mackenzie H, Carter F, Francis NK. First international training and assessment consensus workshop on transanal total mesorectal excision (taTME). Tech Coloproctol. 2016.
161. Curtis NJ, Davids J, Foster JD, Francis NK. Objective assessment of minimally invasive total mesorectal excision performance: a systematic review. Tech Coloproctol. 2017;21(4):259-68.
162. Kennedy RH, Francis EA, Wharton R, Blazeby JM, Quirke P, West NP, et al. Multicenter randomized controlled trial of conventional versus laparoscopic surgery for colorectal cancer within an enhanced recovery programme: EnROL. J Clin Oncol. 2014;32(17):1804-11.
163. Britten N. Qualitative interviews in medical research. BMJ. 1995;311(6999):251-3.

164. Van Spall H, Kassam A, Tollefson TT. Near-misses are an opportunity to improve patient safety: adapting strategies of high reliability organizations to healthcare. Curr Opin Otolaryngol Head Neck Surg. 2015;23(4):292-6.
165. Cuschieri A. Nature of human error: implications for surgical practice. Ann Surg. 2006;244(5):642-8.
166. White AD, Skelton M, Mushtaq F, Pike TW, Mon-Williams M, Lodge JP, et al. Inconsistent reporting of minimally invasive surgery errors. Ann R Coll Surg Engl. 2015;97(8):608-12.
167. Kirwan B. Human error identification in human reliability assessment. Part 1: Overview of approaches. Appl Ergon. 1992;23(5):299-318.
168. Senders DM, NP. Human Error: Cause, prediction and reduction. 1 ed. Hillsdale, NJ: CRC Press; 1991. 168 p.
169. van Rutte P, Nienhuijs SW, Jakimowicz JJ, van Montfort G. Identification of technical errors and hazard zones in sleeve gastrectomy using OCHRA : "OCHRA for sleeve gastrectomy". Surg Endosc. 2017;31(2):561-6.
170. Tang B, Hanna GB, Bax NM, Cuschieri A. Analysis of technical surgical errors during initial experience of laparoscopic pyloromyotomy by a group of Dutch pediatric surgeons. Surg Endosc. 2004;18(12):1716-20.
171. Tang B, Hanna GB, Joice P, Cuschieri A. Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. Arch Surg. 2004;139(11):1215-20.
172. Talebpour M, Alijani A, Hanna GB, Moosa Z, Tang B, Cuschieri A. Proficiency-gain curve for an advanced laparoscopic procedure defined by observation clinical human reliability assessment (OCHRA). Surg Endosc. 2009;23(4):869-75.
173. Cuschieri A. Reducing errors in the operating room: surgical proficiency and quality assurance of execution. Surg Endosc. 2005;19(8):1022-7.
174. Foster JD, Tou S, Curtis NJ, Smart NJ, Acheson A, Maxwell-Armstrong C, et al. Closure of the perineal defect after abdominoperineal excision for rectal adenocarcinoma - ACPGBI Position Statement. Colorectal Dis. 2018;20 Suppl 5:5-23.
175. Shihab OC, Brown G, Daniels IR, Heald RJ, Quirke P, Moran BJ. Patients with low rectal cancer treated by abdominoperineal excision have worse tumors and higher involved margin rates compared with patients treated by anterior resection. Dis Colon Rectum. 2010;53(1):53-6.
176. Wang XT, Li DG, Li L, Kong FB, Pang LM, Mai W. Meta-analysis of oncological outcome after abdominoperineal resection or low anterior resection for lower rectal cancer. Pathol Oncol Res. 2015;21(1):19-27.
177. Marsden MR, Conti JA, Zeidan S, Flashman KG, Khan JS, O'Leary DP, et al. The selective use of splenic flexure mobilization is safe in both laparoscopic and open anterior resections. Colorectal Dis. 2012;14(10):1255-61.
178. Joice P, Hanna GB, Cuschieri A. Ergonomic evaluation of laparoscopic bowel suturing. Am J Surg. 1998;176(4):373-8.
179. Joice P, Hanna GB, Cuschieri A. Errors enacted during endoscopic surgery--a human reliability analysis. Appl Ergon. 1998;29(6):409-14.
180. Wilson PJ. Use of human reliability analysis to evaluate surgical technique for rectal cancer. EThOS depository, British Library: University of Dundee; 2012.
181. Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD, et al. The Clavien-Dindo classification of surgical complications: five-year experience. Ann Surg. 2009;250(2):187-96.

182. Clavien PA, Sanabria JR, Mentha G, Borst F, Buhler L, Roche B, et al. Recent results of elective open cholecystectomy in a North American and a European center. Comparison of complications and risk factors. Ann Surg. 1992;216(6):618-26.
183. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg. 2004;240(2):205-13.
184. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.
185. Loukas C. Video content analysis of surgical procedures. Surg Endosc. 2018;32(2):553-68.
186. Riechelmann RP, Alex A, Cruz L, Bariani GM, Hoff PM. Non-inferiority cancer clinical trials: scope and purposes underlying their design. Ann Oncol. 2013;24(7):1942-7.
187. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. Trials. 2011;12:106.
188. Curtis NJ, Conti JA, Dalton R, Rockall TA, Allison AS, Ockrim JB, et al. 2D versus 3D laparoscopic total mesorectal excision: a developmental multicentre randomised controlled trial. Surg Endosc. 2019.
189. Leung KL, Kwok SP, Lam SC, Lee JF, Yiu RY, Ng SS, et al. Laparoscopic resection of rectosigmoid carcinoma: prospective randomised trial. Lancet. 2004;363(9416):1187-92.
190. Pechlivanides G, Gouvas N, Tsiaoussis J, Tzortzinis A, Tzardi M, Moutafidis M, et al. Lymph node clearance after total mesorectal excision for rectal cancer: laparoscopic versus open approach. Dig Dis. 2007;25(1):94-9.
191. Sakata S, Watson MO, Grove PM, Stevenson AR. The Conflicting Evidence of Three-dimensional Displays in Laparoscopy: A Review of Systems Old and New. Ann Surg. 2016;263(2):234-9.
192. Sorensen SM, Savran MM, Konge L, Bjerrum F. Three-dimensional versus two-dimensional vision in laparoscopy: a systematic review. Surg Endosc. 2016;30(1):11-23.
193. Vettoretto N, Foglia E, Ferrario L, Arezzo A, Cirocchi R, Cocorullo G, et al. Why laparoscopists may opt for three-dimensional view: a summary of the full HTA report on 3D versus 2D laparoscopy by S.I.C.E. (Societa Italiana di Chirurgia Endoscopica e Nuove Tecnologie). Surg Endosc. 2018;32(6):2986-93.
194. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
195. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924-6.
196. Goldet G, Howick J. Understanding GRADE: an introduction. J Evid Based Med. 2013;6(1):50-4.
197. Arezzo A, Vettoretto N, Francis NK, Bonino MA, Curtis NJ, Amparore D, et al. The use of 3D laparoscopic imaging systems in surgery: EAES consensus development conference 2018. Surg Endosc. 2018.
198. Smith R, Day A, Rockall T, Ballard K, Bailey M, Jourdan I. Advanced stereoscopic projection technology significantly improves novice performance of minimally invasive surgical skills. Surg Endosc. 2012;26(6):1522-7.
199. Smith R, Schwab K, Day A, Rockall T, Ballard K, Bailey M, et al. Effect of passive polarizing three-dimensional displays on surgical performance for experienced laparoscopic surgeons. Br J Surg. 2014;101(11):1453-9.

200. Wilhelm D, Reiser S, Kohn N, Witte M, Leiner U, Muhlbach L, et al. Comparative evaluation of HD 2D/3D laparoscopic monitors and benchmarking to a theoretically ideal 3D pseudodisplay: even well-experienced laparoscopists perform better with 3D. Surg Endosc. 2014;28(8):2387-97.
201. Curro G, Lazzara S, La Malfa G, Giovanni P, De Leo E, Fortugno A, et al. Three dimensional (3D) versus two-dimensional (2D) laparoscopic oncological colorecta surgery: A single-surgeon perspective randomized comparative study. Eur J Surg Oncol. 2016;42(10):S206.
202. Curro G, Cogliandolo A, Bartolotta M, Navarra G. Three-Dimensional Versus Two-Dimensional Laparoscopic Right Hemicolectomy. J Laparoendosc Adv Surg Tech A. 2016;26(3):213-7.
203. Dias RD, Ngo-Howard MC, Boskovski MT, Zenati MA, Yule SJ. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. Br J Surg. 2018;105(5):491-501.
204. Francis NK, Curtis NJ, Crilly L, Noble E, Dyke T, Hipkiss R, et al. Does the number of operating specialists influence the conversion rate and outcomes after laparoscopic colorectal cancer surgery? Surg Endosc. 2018;32(8):3652-8.
205. Curtis NJ, Noble E, Salib E, Hipkiss R, Meachim E, Dalton R, et al. Does hospital readmission following colorectal cancer resection and enhanced recovery after surgery affect long term survival? Colorectal Dis. 2017;19(8):723-30.
206. Billingham SA, Whitehead AL, Julious SA. An audit of sample sizes for pilot and feasibility trials being undertaken in the United Kingdom registered in the United Kingdom Clinical Research Network database. BMC Med Res Methodol. 2013;13:104.
207. Ljungqvist O, Scott M, Fearon KC. Enhanced Recovery After Surgery: A Review. JAMA Surg. 2017;152(3):292-8.
208. SG H. NASA-Task Load index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society

Ergonomics Society Annual Meeting 2006; 1/1/20062006. p. 904-8.
209. SG H, LE S. Development of NASA-TLX (Task load index): Results of empirical and theoretical research. Advances in Psychology. 1998;52:139-83.
210. Curtis NJ, Thomas C, Dennison G, Ockrim J, Conti JA, Dalton R, et al. Factors predicting operative difficulty of laparoscopic total mesorectal excision. Dis Colon Rectum. 2019.
211. Escal L, Nougaret S, Guiu B, Bertrand MM, de Forges H, Tetreau R, et al. MRI-based score to predict surgical difficulty in patients with rectal cancer. Br J Surg. 2018;105(1):140-6.
212. Schwab KE, Curtis NJ, Whyte MB, Smith RV, Rockall TA, Ballard K, et al. 3D laparoscopy does not reduce operative duration or errors in day-case laparoscopic cholecystectomy: a randomised controlled trial. Surg Endosc. 2019.
213. West NP, Finan PJ, Anderin C, Lindholm J, Holm T, Quirke P. Evidence of the oncologic superiority of cylindrical abdominoperineal excision for low rectal cancer. J Clin Oncol. 2008;26(21):3517-22.
214. West NP, Hohenberger W, Weber K, Perrakis A, Finan PJ, Quirke P. Complete mesocolic excision with central vascular ligation produces an oncologically superior specimen compared with standard surgery for carcinoma of the colon. J Clin Oncol. 2010;28(2):272-8.
215. Hanna GB, Shimi SM, Cuschieri A. Randomised study of influence of two-dimensional versus three-dimensional imaging on performance of laparoscopic cholecystectomy. Lancet. 1998;351(9098):248-51.

216. Hirst A, Philippou Y, Blazeby J, Campbell B, Campbell M, Feinberg J, et al. No Surgical Innovation Without Evaluation: Evolution and Further Development of the IDEAL Framework and Recommendations. Ann Surg. 2018.
217. Blencowe NS, Mills N, Cook JA, Donovan JL, Rogers CA, Whiting P, et al. Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. Br J Surg. 2016;103(10):1377-84.
218. Park JS, Choi GS, Park SY, Kim HJ, Ryuk JP. Randomized clinical trial of robot-assisted versus standard laparoscopic right colectomy. Br J Surg. 2012;99(9):1219-26.
219. Criss CN, MacEachern MP, Matusko N, Dimick JB, Maggard-Gibbons M, Gadepalli SK. The Impact of Corporate Payments on Robotic Surgery Research: A Systematic Review. Ann Surg. 2019;269(3):389-96.
220. Gjeraa K, Spanager L, Konge L, Petersen RH, Ostergaard D. Non-technical skills in minimally invasive surgery teams: a systematic review. Surg Endosc. 2016;30(12):5185-99.
221. Bonrath EM, Dedy NJ, Zevin B, Grantcharov TP. Defining technical errors in laparoscopic surgery: a systematic review. Surg Endosc. 2013;27(8):2678-91.
222. Bonrath EM, Zevin B, Dedy NJ, Grantcharov TP. Error rating tool to identify and analyse technical errors and events in laparoscopic surgery. Br J Surg. 2013;100(8):1080-8.
223. Hanna GB. Influence of imaging on endoscopic task performance: University of Dundee; 1998.
224. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Med Teach. 2012;34(11):960-92.
225. Markar SR, Wiggins T, Ni M, Steyerberg EW, Van Lanschot JJ, Sasako M, et al. Assessment of the quality of surgery within randomised controlled trials for the treatment of gastro-oesophageal cancer: a systematic review. Lancet Oncol. 2015;16(1):e23-31.
226. Monson JR, Probst CP, Wexner SD, Remzi FH, Fleshman JW, Garcia-Aguilar J, et al. Failure of evidence-based cancer care in the United States: the association between rectal cancer treatment, cancer center volume, and geography. Ann Surg. 2014;260(4):625-31; discussion 31-2.
227. Mishra A, Catchpole K, Dale T, McCulloch P. The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. Surg Endosc. 2008;22(1):68-73.
228. Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RS. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). World J Surg. 2011;35(9):1961-9.
229. Nickel F, Hendrie JD, Stock C, Salama M, Preukschas AA, Senft JD, et al. Direct Observation versus Endoscopic Video Recording-Based Rating with the Objective Structured Assessment of Technical Skills for Training of Laparoscopic Cholecystectomy. Eur Surg Res. 2016;57(1-2):1-9.
230. Francis NK, Curtis NJ, Conti JA, Foster JD, Bonjer HJ, Hanna GB, et al. EAES classification of intraoperative adverse events in laparoscopic surgery. Surg Endosc. 2018.
231. Young J, Badgery-Parker T, Dobbins T, Jorgensen M, Gibbs P, Faragher I, et al. Comparison of ECOG/WHO performance status and ASA score as a measure of functional status. J Pain Symptom Manage. 2015;49(2):258-64.
232. Roth D, Pace NL, Lee A, Hovhannisyan K, Warenits AM, Arrich J, et al. Airway physical examination tests for detection of difficult airway management in apparently normal adult patients. Cochrane Database Syst Rev. 2018;5:CD008874.
233. Ubbink DT, Visser A, Gouma DJ, Goslings JC. Registration of surgical adverse outcomes: a reliability study in a university hospital. BMJ Open. 2012;2(3).

234. Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. Health Technol Assess. 2001;5(22):1-194.
235. Keller DS, Delaney CP, Senagore AJ, Feldman LS, Force SST. Uptake of enhanced recovery practices by SAGES members: a survey. Surg Endosc. 2017;31(9):3519-26.
236. Dabbas N, Adams K, Chave H, Branagan G. Current practice in abdominoperineal resection: an email survey of the membership of the Association of Coloproctology. Ann R Coll Surg Engl. 2012;94(3):173-6.
237. Mirnezami A, Mirnezami R, Chandrakumaran K, Sasapu K, Sagar P, Finan P. Increased local recurrence and reduced survival from colorectal cancer following anastomotic leak: systematic review and meta-analysis. Ann Surg. 2011;253(5):890-9.
238. Curtis NJ, Dennison G, Brown CSB, Hewett PJ, Hanna GB, Stevenson ARL, et al. Clinical Evaluation of Intraoperative Near Misses in Laparoscopic Rectal Cancer Surgery. Ann Surg. 2019.
239. Mathis KL, Cima RR. Quality assurance in colon and rectal cancer surgery. Surg Oncol Clin N Am. 2014;23(1):11-23.
240. Sarker SK, Chang A, Vincent C, Darzi AW. Technical skills errors in laparoscopic cholecystectomy by expert surgeons. Surg Endosc. 2005;19(6):832-5.
241. Heinrich HW. Industrial Accident Prevention: a scientific approach. 1 ed. New York: McGraw-Hill; 1931.
242. Yorio PL, Moore SM. Examining Factors that Influence the Existence of Heinrich's Safety Triangle Using Site-Specific H&S Data from More than 25,000 Establishments. Risk Anal. 2018;38(4):839-52.
243. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. BMJ. 2015;351:h4672.
244. Henken KR, Jansen FW, Klein J, Stassen LP, Dankelman J, van den Dobbelsteen JJ. Implications of the law on video recording in clinical practice. Surg Endosc. 2012;26(10):2909-16.
245. Howell AM, Burns EM, Bouras G, Donaldson LJ, Athanasiou T, Darzi A. Can Patient Safety Incident Reports Be Used to Compare Hospital Safety? Results from a Quantitative Analysis of the English National Reporting and Learning System Data. PLoS One. 2015;10(12):e0144107.
246. Kuhnen AH, Marcello PW, Roberts PL, Read TE, Schoetz DJ, Rusin LC, et al. Can the national surgical quality improvement program provide surgeon-specific outcomes? Dis Colon Rectum. 2015;58(2):247-53.
247. Foster JD, Francis NK. Objective assessment of technique in laparoscopic colorectal surgery: what are the existing tools? Tech Coloproctol. 2015;19(1):1-4.
248. Blencowe NS, Boddy AP, Harris A, Hanna T, Whiting P, Cook JA, et al. Systematic review of intervention design and delivery in pragmatic and explanatory surgical randomized clinical trials. Br J Surg. 2015;102(9):1037-47.
249. Tudur Smith C, Marcucci M, Nolan SJ, Iorio A, Sudell M, Riley R, et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. Cochrane Database Syst Rev. 2016;9:MR000007.
250. Bunt AM, Hermans J, Boon MC, van de Velde CJ, Sasako M, Fleuren GJ, et al. Evaluation of the extent of lymphadenectomy in a randomized trial of Western- versus Japanese-type surgery in gastric cancer. J Clin Oncol. 1994;12(2):417-22.
251. Makary MA. The power of video recording: taking quality to the next level. JAMA. 2013;309(15):1591-2.

252. Bezemer J, Cope A, Korkiakangas T, Kress G, Murtagh G, Weldon SM, et al. Microanalysis of video from the operating room: an underused approach to patient safety research. BMJ Qual Saf. 2017;26(7):583-7.
253. Deijen CL, Velthuis S, Tsai A, Mavroveli S, ES dL-dK, Sietses C, et al. COLOR III: a multicentre randomised clinical trial comparing transanal TME versus laparoscopic TME for mid and low rectal cancer. Surg Endosc. 2016;30(8):3210-5.
254. Moran B, Karandikar S, Geh I. Association of Coloproctology of Great Britain & Ireland (ACPGBI): Guidelines for the Management of Cancer of the Colon, Rectum and Anus (2017) - Introduction. Colorectal Dis. 2017;19 Suppl 1:6-8.
255. Gooiker GA, Lemmens VE, Besselink MG, Busch OR, Bonsing BA, Molenaar IQ, et al. Impact of centralization of pancreatic cancer surgery on resection rates and survival. Br J Surg. 2014;101(8):1000-5.
256. O'Mahoney PRA, Yeo HL, Sedrakyan A, Trencheva K, Mao J, Isaacs AJ, et al. Centralization of pancreatoduodenectomy a decade later: Impact of the volume-outcome relationship. Surgery. 2016;159(6):1528-38.
257. Ahola R, Siiki A, Vasama K, Vornanen M, Sand J, Laukkarinen J. Effect of centralization on long-term survival after resection of pancreatic ductal adenocarcinoma. Br J Surg. 2017;104(11):1532-8.
258. Aquina CT, Probst CP, Becerra AZ, Iannuzzi JC, Kelly KN, Hensley BJ, et al. High volume improves outcomes: The argument for centralization of rectal cancer surgery. Surgery. 2016;159(3):736-48.

Appendix 1 – Permissions for reproduction of previously published figures and tables

Written permissions from copyright holders for the reproduction of previously published figures and tables are presented on the next two pages. Each are acknowledged in the respective figure or table.

Please note, my Copyright Clearance Center RightsLink account was created during previous employment at the University of Southampton.



SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Mar 20, 2019

This Agreement between Nathan J Curtis ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4552470458194
License date	Mar 19, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Anatomy of the Colon, Rectum, Anus, and Pelvic Floor
Licensed Content Author	Thilo Wedel
Licensed Content Date	Jan 1, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery
Institution name	n/a
Expected presentation date	Apr 2019
Portions	Figure 2.2b Figure 2.6
Requestor Location	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, Hampshire SO16 6YD United Kingdom Attn: Nathan J Curtis
Total	0.00 GBP

Subject Re: Reproduction of virtual library slide
From medpathw <pathology-webmaster@leeds.ac.uk>
To nathancurtis@doctors.org.uk <nathancurtis@doctors.org.uk>, danrentmaor@nhs.net <danrentmaor@nhs.net>
Date 2019-03-19 08:34

Nathan

Yes it is possible - and you can have permisson to do so, but please cite / reference "University of Leeds" and our website www.virtualpathology.leeds.ac.uk
If you want some advice on how to get an image suitable for publication then please contact me on the email address below.

Best wishes for your thesis

Martin Waterhouse

4.11 Wellcome Trust Brenner Building
St. James's University Hospital
Leeds
LS9 7TF

m.c.waterhouse@leeds.ac.uk

+44(0)113 3438929

From: nathancurtis@doctors.org.uk <nathancurtis@doctors.org.uk>
Sent: 18 March 2019 16:27:24
To: medpathw
Subject: Reproduction of virtual library slide

Hello,
Is it possible to reproduce a rectal cancer slide from your amazing collection in my thesis? It will eventually be publically accessible through the Imperial Spiral library http://spiral.imperial.ac.uk:8080/ but there is no commercial aspect. Of course, your library would be fully acknowledged.
Many thanks,
Nathan Curtis
PhD student, Imperial College London

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Mar 20, 2019

This Agreement between Nathan J Curtis ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4544361040930
License date	Mar 08, 2019
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	British Journal of Surgery
Licensed Content Title	The mesorectum in rectal cancer surgery—the clue to pelvic recurrence?
Licensed Content Author	R. J. Heald, E. M. Husband, R. D. H. Ryall
Licensed Content Date	Dec 7, 2005
Licensed Content Volume	69
Licensed Content Issue	10
Licensed Content Pages	4
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1
Will you be translating?	No
Title of your thesis / dissertation	Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery
Expected completion date	Apr 2019
Expected size (number of pages)	280
Requestor Location	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, Hampshire SO16 6YD United Kingdom Attn: Nathan J Curtis
Publisher Tax ID	EU826007151
Total	0.00 GBP

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Mar 20, 2019

This Agreement between Nathan J Curtis ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4544370182413
License date	Mar 08, 2019
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Colorectal Disease
Licensed Content Title	Scoring the quality of total mesorectal excision for the prediction of cancer-specific outcome
Licensed Content Author	D. Leonard, F. Penninckx, A. Laenen, et al
Licensed Content Date	Apr 21, 2015
Licensed Content Volume	17
Licensed Content Issue	5
Licensed Content Pages	8
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1
Will you be translating?	No
Title of your thesis / dissertation	Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery
Expected completion date	Apr 2019
Expected size (number of pages)	280
Requestor Location	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, Hampshire SO16 6YD United Kingdom Attn: Nathan J Curtis
Publisher Tax ID	EU826007151
Total	0.00 GBP

Clockwise from top left: Wedel et al (figure 1.1), Virtual pathology slide library University of Leeds (in further communication they confirmed email approval is their usual policy for reproduction. Figure 1.2), Heald et al (figure 1.4) and Leonard et al (figure 1.6).



WOLTERS KLUWER HEALTH, INC. LICENSE TERMS AND CONDITIONS

Mar 20, 2019

This Agreement between Nathan J Curtis ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	4544370428737
License date	Mar 08, 2019
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	Annals of Surgery
Licensed Content Title	Methods of Quality Assurance in Multicenter Trials in Laparoscopic Colorectal Surgery: A Systematic Review.
Licensed Content Author	Foster, Jake; Mackenzie, Hugh; Nelson, Heidi; Hanna, George; Francis, Nader
Licensed Content Date	Aug 1, 2014
Licensed Content Volume	260
Licensed Content Issue	2
Type of Use	Dissertation/Thesis
Requestor type	Individual
STM publisher name	
Portion	Figures/table/illustration
Number of figures/tables/illustrations	1
Figures/tables/illustrations used	Table 6
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery
Expected completion date	Apr 2019
Estimated size(pages)	280
Requestor Location	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, Hampshire SO16 6YD United Kingdom Attn: Nathan J Curtis
Publisher Tax ID	EU826013006
Billing Type	Invoice
Billing Address	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, United Kingdom SO16 6YD Attn: Nathan J Curtis
Total	0.00 GBP

WOLTERS KLUWER HEALTH, INC. LICENSE TERMS AND CONDITIONS

Mar 20, 2019

This Agreement between Nathan J Curtis ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	4544370765318
License date	Mar 08, 2019
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	Annals of Surgery
Licensed Content Title	Is Competency Assessment at the Specialist Level Achievable? A Study for the National Training Programme in Laparoscopic Colorectal Surgery in England.
Licensed Content Author	Miskovic, Danilo; MD, PhD; Ni, Melody; Wyles, Susannah; MBBS, MSc; Kennedy, Robin; MD, FRCS; Francis, Nader; PhD, FRCS; Parvaiz, Arnjad; FRCS, FRCS; Cunningham, Chris; MD, FRCS; Rockall, Timothy; MD, FRCS; Gudgeon, Andrew; MBBS, FRCS; Coleman, Mark; MD, FRCS; Hanna, George; PhD, FRCS
Licensed Content Date	Mar 1, 2013
Licensed Content Volume	257
Licensed Content Issue	3
Type of Use	Dissertation/Thesis
Requestor type	Individual
STM publisher name	
Portion	Figures/table/illustration
Number of figures/tables/illustrations	1
Figures/tables/illustrations used	Figure 1
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery
Expected completion date	Apr 2019
Estimated size(pages)	280
Requestor Location	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, Hampshire SO16 6YD United Kingdom Attn: Nathan J Curtis
Publisher Tax ID	EU826013006
Billing Type	Invoice
Billing Address	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, United Kingdom SO16 6YD Attn: Nathan J Curtis
Total	0.00 GBP

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Mar 20, 2019

This Agreement between Nathan J Curtis ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4544381269834
License date	Mar 08, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Techniques in Coloproctology
Licensed Content Title	Application of objective clinical human reliability analysis (OCHRA) in assessment of technical performance in laparoscopic rectal cancer surgery
Licensed Content Author	J. D. Foster, D. Miskovic, A. S. Allison et al
Licensed Content Date	Jan 1, 2016
Licensed Content Volume	20
Licensed Content Issue	6
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery
Institution name	n/a
Expected presentation date	Apr 2019
Portions	Figure 1
Requestor Location	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, Hampshire SO16 6YD United Kingdom Attn: Nathan J Curtis
Billing Type	Invoice
Billing Address	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, United Kingdom SO16 6YD Attn: Nathan J Curtis
Total	0.00 GBP

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Mar 20, 2019

This Agreement between Nathan J Curtis ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4544380385682
License date	Mar 08, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Surgical Endoscopy
Licensed Content Title	EAES classification of intraoperative adverse events in laparoscopic surgery
Licensed Content Author	N. K. Francis, N. J. Curtis, J. A. Conti et al
Licensed Content Date	Jan 1, 2018
Licensed Content Volume	32
Licensed Content Issue	9
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	3
Will you be translating?	no
Circulation/distribution	>50.000
Author of this Springer Nature content	yes
Title	Application of intraoperative quality assurance to laparoscopic total mesorectal excision surgery
Institution name	n/a
Expected presentation date	Apr 2019
Portions	Table 1 Figure 1 and 2
Requestor Location	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, Hampshire SO16 6YD United Kingdom Attn: Nathan J Curtis
Billing Type	Invoice
Billing Address	Nathan J Curtis Cancer Sciences Division Somers Cancer Sciences Building, MP824 Southampton General Hospital Southampton, United Kingdom SO16 6YD Attn: Nathan J Curtis
Total	0.00 GBP

Clockwise from top left: Foster et al (table 1.), Miskovic et al (figure 1.8), Foster et al (figure 3.1) and Francis et al figures 6.1, 6.2 and table 6.1).

Appendix 2 – TMEdt individual case score quartile analysis

This table displays outcome data for cases grouped into quartiles based upon TMEdt score irrespective of the surgeon performing the case. Overall the data is comparable with that seen for the surgeon quartile analysis (table 5.3). No differences were seen in the oncological outcome data (not shown).

		TMEdt score case quartiles									Three groups	Upper-lower
		Lower			Interquartile range			Upper			p	p
		Median	IQR	%	Median	IQR	%	Median	IQR	%		
Performance measurement												
Laparoscopic TME performance tool score		34	31-36		40	39-42		46	45-46		**<0.001**	**<0.001**
Pelvic OCHRA error count		7	5-11		6	4-8		3	2-5		**<0.001**	**<0.001**
Demographics												
Age (years)		68	61-74		66	59-75		65	55-73		0.502	0.269
Body Mass Index (kg/m2)		28	24-30		28	25-30		24	22-27		**<0.001**	**<0.001**
Tumour Height (cm)		8.0	6.0-10.0		7.5	5-10		8.0	7-12		0.084	0.555
Sex	Female			24.5%			43.8%			52.4%	**0.015**	**0.005**
	Male			75.5%			56.3%			47.6%		
Neoadjuvant Treatment				32.1%			43.8%			38.1%	0.4	0.543
Tumour stage (Pathologically defined)	PathCR			1.9%			1.3%			0.0%	0.772	0.946
	I			26.4%			23.8%			38.1%		
	II			28.3%			28.8%			14.3%		
	III			43.4%			42.5%			45.2%		
	IV			0.0%			3.8%			2.4%		
Operative data												
Operative duration (minutes)		281	240-335		243	195-278		190	160-220		**<0.001**	**<0.001**
Estimated blood loss (ml)		200	90-300		100	50-150		50	20-60		**<0.001**	**<0.001**
Histopathological outcomes												
	Mesorectal Fascia			76.0%			83.3%			97.4%	**0.026**	**0.006**

Plane of mesorectal dissection	Intramesorectal			18.0%			10.3%			0.0%		
	Muscularis Propria			6.0%			6.4%			2.6%		
ALaCART Composite Endpoint Success				75.5%			81.3%			92.9%	0.085	**0.025**
Lymph node yield		17	12-25		16	12-21		19	14-25		0.219	0.384
Clinical Outcomes												
Any 30 day morbidity event				54.7%			50.0%			38.1%	0.26	0.109
Number of 30 day morbidity events	None			45.3%			50.0%			61.9%	0.125	**0.049**
	1			20.8%			21.3%			19.0%		
	2			9.4%			20.0%			11.9%		
	3			13.2%			6.3%			4.8%		
	4			7.5%			2.5%			2.4%		
	5			3.8%			0.0%			0.0%		
Highest grade morbidity experienced (Common terminology for adverse event reporting system)	None			45.3%			50.0%			61.9%	0.17	0.066
	1			9.4%			8.8%			7.1%		
	2			24.5%			27.5%			23.8%		
	3			13.2%			12.5%			4.8%		
	4			7.5%			1.3%			2.4%		
Unplanned Re-operation				9.4%			5.0%			2.4%	0.318	0.163
Anastomotic Leak				7.5%			5.0%			0.0%	0.212	0.07
Length of stay (days)		9	6-14		8	6-14		7	5-9		**0.023**	**0.005**
Hospital Readmission				31.8%			19.0%			7.7%	0.221	0.106

Quartiles based upon individual case TMEdt score rather than the surgeon performing the procedure. Overall a similar significant pattern in seen with excellent outcomes following high scoring surgery.

Appendix 3 – Full laparoscopic TME performance tool data

87 laparoscopic TME case videos were seen to capture all resection and anastomosis tasks allowing full application of the L-TMEpt (14 ALaCaRT and 73 2D3D). Demographic and tumour details were unchanged compared to the TMEdt analyses. Mean age, body mass index and tumour height from the anal verge was 66 (±11), 27kg/m^2 (±5) and 8.5cm (±3cm) respectively. 52% were males and 32% received neoadjuvant chemoradiotherapy.

OCHRA analysis identified 698 pelvic errors (median 7, mode 7, IQR 4-11, range 0-31). Moderate negative correlation was seen with L-TMEpt scores (r^s=-0.669, p<0.001). Substantial variation in L-TMEpt scores are observed (median 53, interquartile range 49-58, range 35-64). Case complexity grades were as: grade 1 40 (46%), grade 2 36 (41.4%) and grade 3 11 (12.6%). A similar relationship with case grades was seen (55 (52-59) vs. 52 (49-56) vs. 48 (45-51), p=0.001). Operative duration and blood loss were both negatively correlated with tool scores (r^s=-0.466, p<0.001 and r^s=-0.457, p<0.001 respectively).

No differences were seen in histopathologically assessed as mesorectal fascial plane surgery (53 (49-56) vs. 53 (50-58), p=0.624) nor those successfully meeting the ALaCaRT composite endpoint (53 (49-55) vs. 53 (50-58), p=0.587). There was no relationship between lymph node yield and tool scores (r^s=-0.084, p=0.438).

46 (53%) of patients developed a morbidity event within 30 days of surgery. Higher L-TMEpt scores were seen in those with uneventful recoveries (55 (52-58) vs. 52 (47-56), p=0.02). There was no difference in L-TMEpt scores in patients that leaked (n=5, 5.7%, 47 (46-52) vs. 54 (50-58), p=0.075), underwent unplanned re-operation (n=5, 5.7%, 52 (46-53) vs. 53 (50-58) p=0.459) or readmitted to hospital (19.2%, 51 (45-52) vs. 53 (49-56), p=0.051). Tool scores were not correlated with length of stay (r^s=-0.103, p=0.342).

16 credentialed surgeons performed the 87 cases with substantial variation in their personal mean score (median 53, range 49-63). Surgeon quartiles were applied with data shown in table 5.6.

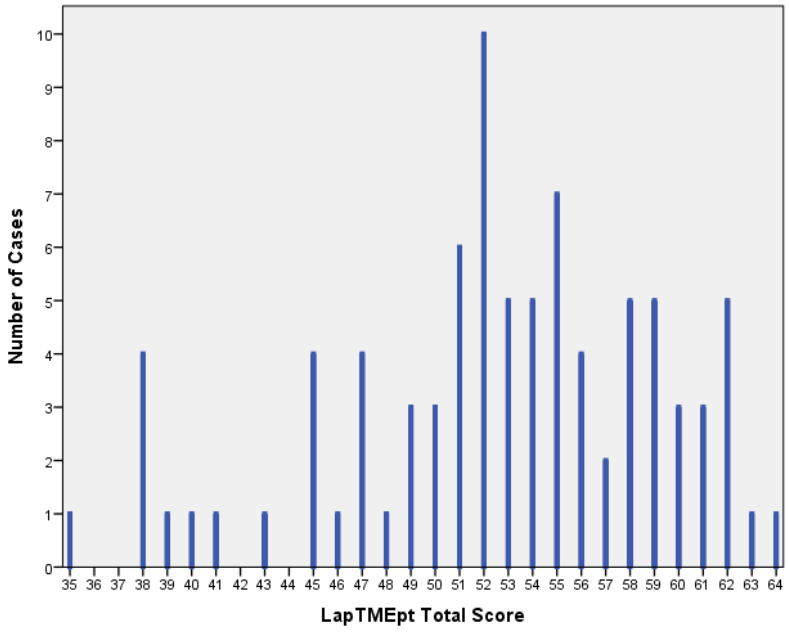
Similar results were identified as there was significantly more morbidity in the lower quartile (78.9% vs. 43.5%, p=0.021). A similar pattern of clinically relevant but not statistically significant reductions in anastomotic leak (6.5% vs. 0; p=0.161), length of stay (9 (5-14) vs. 7 (5-11) days, p=0.2) and hospital readmission (33% vs. 0, p=0.186) were seen.

There were no differences in any of the oncological outcomes shown in table 5.4 (not shown).

		L-TMEpt Surgeon Quartiles									Three groups	Upper-lower
		Lower			Interquartile			Upper			p	p
		Median	IQR	%	Median	IQR	%	Median	IQR	%		
Performance measurement												
L-TME performance tool score (TMEdt)		45	38-47		53	51-55		60	59-62		**<0.001**	**<0.001**
Pelvic error count (OCHRA)		12	10-18		7	5-9		3	2-6		**<0.001**	**<0.001**
Demographics												
Age (years)		71	61-76		65	59-72		65	54-75		0.383	0.312
Body Mass Index (kg/m2)		27	23-30		28	25-30		26	23-28		0.168	0.336
Tumour Height (cm)		8	7-11		8	6-10		9	7-12		0.221	0.646
Sex	Female			26.3%			55.6%			52.2%	0.095	0.093
	Male			73.7%			44.4%			47.8%		
Neoadjuvant Treatment				10.5%			37.8%			39.1%	0.075	**0.038**
Tumour stage (Histopathologically defined)	PathCR			0			4.4%			0	0.727	0.516
	I			36.8%			26.7%			38.1%		
	II			26.3%			37.8%			21.7%		
	III			36.8%			28.9%			30.4%		
	IV			0			2.2%			0		
Operative data												
Operative duration (minutes)		310	270-380		270	240-310		207	175-270		**<0.001**	**<0.001**
Estimated blood loss (ml)		100	80-200		60	40-100		50	20-100		**0.002**	**<0.001**

Histopathological outcomes												
Plane of mesorectal dissection	Mesorectal Fascia			88.2%			83.3%			90.5%		
	Intramesorectal			11.8%			9.5%			0.0%	0.740	0.912
	Muscularis Propria			0			7.1%			9.5%		
ALaCART Composite Endpoint Success				89.5%			88.2%			91.3%	0.533	0.842
Lymph node yield		20	14-31		18	15-23		18	12-26		0.572	0.356
Clinical Outcomes												
Any 30 day morbidity event				78.9%			46.7%			43.5%	**0.037**	**0.021**
Number of 30 day morbidity events	None			21.1%			53.3%			56.5%		
	1			26.3%			20%			8.7%		
	2			21.1%			17.8%			21.7%	**0.026**	**0.041**
	3			21.1%			4.4%			8.7%		
	4			0			4.4%			4.3%		
	5			11.5%			0			0		
Highest grade morbidity experienced (Common terminology for adverse event reporting system)	None			21.1%			53.3%			56.5%		
	1			21.1%			8.9%			0		
	2			31.6%			26.7%			34.8%	0.071	0.061
	3			21.1%			6.7%			4.3%		
	4			5.3%			4.4%			4.3%		
Unplanned Re-operation				10.5%			4.4%			4.3%	0.602	0.445
Anastomotic Leak				15.8%			4.4%			0	0.081	0.051
Length of stay (days)		12	5-22		6	8-13		7	5-18		0.653	0.417
Hospital Readmission				26.3%			19.5%			7.7%	0.425	0.192

Surgeon quartile analysis using full L-TMEpt (four column) data.



Full L-TMEpt scores and distribution. A similar, although wider spread is seen.