The Multi-Institutional Validation and Assessment of Training Modalities in Robotic Surgery
The MARS Project

Raison, Nicholas

Awarding institution: King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

• Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
• Non Commercial: You may not use this work for commercial purposes.
• No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
The Multi-Institutional Validation and Assessment of Training Modalities in Robotic Surgery: The MARS Project

Nicholas Tobias Johannes Raison

Supervisor: Prof. P Dasgupta
Dr K Ahmed

Faculty of Life Sciences & Medicine
King’s College London

This dissertation is submitted for the degree of

Doctor of Philosophy

December 2020
Abstract

Surgical simulation is increasingly being recognised for the considerable benefits it can confer to medical training and assessment. Simulation training has been applied widely to the majority of medical and surgical disciplines and is supported by a large body of evidence. Robotic surgery, as an emerging speciality, has been enthusiastically adopted by both surgeons and patients. To ensure the safe and successful implementation of this new technology, evidence-based and effective training programmes need to be developed. Robotic surgery offers the unique opportunity to embed the principles of simulation based education within the new training programmes. The aim of this work, entitled the Multi-Institutional Validation and Assessment of Training Modalities in Robotic Surgery (The MARS) Project is to identify, assess and validate the key components required for the delivery of effective robotic surgical simulation.

Each potential component of the robotic simulation curriculum was addressed separately. Initially a comprehensive review of current simulation practices was conducted to inform the rest of the work. Uniquely the review focussed on the process and structure of simulation training as reported in the published literature in contrast to the outcomes based reviews undertaken previously. All the principle modalities of simulation training applicable to robotic surgery and their potential role within a robotic surgical curriculum were then addressed. The role of virtual reality (VR) training was explored for both basic and procedural training. VR is widely used for basic surgical skills training but this training remains largely
ad hoc. To date no structured VR simulation training programmes have been developed for robotic surgery. Deliberate practice is widely be recognised for its important role in training, greatly enhancing skill acquisition. Incorporating these principles, I devised a robotic VR basic skills training programme employing formal benchmark scores. Procedural VR simulation represents an emerging simulation modality that offers the opportunity for more advanced skills training. As part of this work, I undertook the first validation assessment of procedural VR training for robotic surgery. Dry lab training is another common approach to basic skills training. In comparison to other surgical techniques, dry lab training is less commonly used in robotic surgery. To address this I completed a comparative analysis of dry lab and basic VR training to determine whether either technique offers any advantages. The key modalities for advanced, procedural surgical skills training are human cadaver and live animal wet lab training. Yet neither form of simulation have been comprehensively evaluated for robotic surgery. Studies of the role of both modalities for robotic training were undertaken. Given the absence of human cadaveric training for robotic training, I developed the first UK training programme and evaluated the outcomes from five courses run between 2015 and 2017. To assess the role of live animal training, I conducted a survey of outcomes following training at the Minimal Invasiv Udviklings Center, Aalborg University Hospital, Denmark. Alongside technical skills training, non technical skills (NTS) are of critical importance to all surgical training. As part of this work, I created the first NTS rating system specifically for robotic surgery. Finally mental imagery (MI) training, widely used in competitive sports as an adjunct to training, remains under-utilised in surgery. Prior studies have demonstrated that potential for MI in surgery. I developed a MI programme for robotic surgery and completed an RCT evaluating its role for basic technical and non technical skills training. The results and conclusions from these studies supporter further recommendations on
a structure for simulation-based robotic surgical training. Further work will be required to validate the effectiveness of the training curriculum.
ACKNOWLEDGEMENTS

First and foremost, I would like to thank Prof Prokar Dasgupta, my mentor and primary supervisor for his unwavering support and guidance. I will always be grateful for the opportunities he has given me and his tireless encouragement. I owe my academic career to him and will always be thankful. I am also entirely grateful to Kamran Ahmed, who has been so instrumental in making my thesis a success. The MARS project would not have been possible without the kind and generous grant from the Vattikuti foundation and I am indebted to them.

I would like to thank my colleague and friend, Abdullatif Aydin, for all his help and assistance he has given me and the MARS project over the years. Completing this thesis has allowed me to make some great collaborative links. I would especially like to mention Taka Abe, whom I was lucky enough to work with during his fellowship year. I will always be grateful to Prof Shamim Khan both for his help and support for my thesis as well as to my clinical training. Whilst completing the MARS project, I have been lucky enough to work with some incredible surgeons and educators from around the world. I would like to thank Ben Challacombe, Prof Henk Van Der Poel, Prof Naeem Soomro, Senthil Nathan, Andrea Gavazzi, Prof Alex Mottrie, Christian Brown, Craig McIlhenny, Giacomo Novara and Nicolo Buffi. I would also like to thank all my students with whom I have worked with over the years
The MARS project would also not have been possible without the many participants who kindly gave up their time. I would like to thank them all for their enthusiasm and willing support.

Above all I will always be thankful to my wife, Shreya, for her incredible and patient support. I could not have done it without her. Equally, I am grateful to my parents and parents-in-law for their continued love and support. Finally, my darling daughter Apsara for entertaining me through this process.
Abbreviations

ANCOVA    Analysis of Covariance
ANOVA     Analysis of Variance
AUA       American Urological Association
BND       Bladder Neck Dissection
DVSS      Da Vinci Surgical Simulator
EAU       European Association of Urology
ERUS      EAU Robotic Urology Section
FLS       Fundamentals of Laparoscopic Surgery[230]
FSRS      Fundamentals of Robotic Surgery[232]
GEARS     Global Evaluation Assessment of Robot Skills[85]
GRS       Global Rating Scale
HOT       Hands on Training
ICARS     Interpersonal and Cognitive Assessment for Robotic Surgery[193]
ICC       Intraclass Correlation Coefficient
MI        Motor Imagery
MIS       Minimally Invasive Surgery
MIQ       Movement Imagery Questionnaire
NTS       Non-Technical Skills
NOTSS     Non-Technical Skills for Surgeons[269]
OR  Operating Room
RARP  Radical Robot Assisted Prostatectomy
RACE  Robotic Anastomosis Competency Evaluation[203]
RoSS  Robotic Surgery Simulator
SD  Standard Deviation
SDL  Self Directed Learning
UVA  Urethrovvesical Anastomosis
VAS  Visual Analogue Scale
VR  Virtual Reality
WCE  World Congress of Endourology
# Table of Contents

List of figures xvi

List of tables xx

1 Simulation Training in Surgical Education 1
   1.1 Overview and Structure of the Thesis ...................... 1
   1.2 A Brief History of Simulation Training in Surgery .......... 4
   1.3 Robot Assisted Laparoscopic Surgery ..................... 8
   1.4 Simulation Training in Surgery ........................... 9
      1.4.1 Virtual Reality Simulation Training .................. 10
      1.4.2 Dry Lab Simulation Training .......................... 14
      1.4.3 Wet Lab Simulation Training .......................... 16
      1.4.4 Motor Imagery Training ............................... 16
      1.4.5 Non-Technical Skills Training ........................ 17
      1.4.6 The Effectiveness of Surgical Simulation Training .... 18
      1.4.7 Validation Assessment for Simulation Training ........ 20
      1.4.8 Implementation of Simulation into Surgical Training ... 22
   1.5 Aims of the MARS Project ................................ 24

2 The State of Training in Surgery: A Systematic Review 25
   2.1 Introduction ............................................. 25
   2.2 Methods .................................................. 26
### Table of contents

2.2.1 Aims ......................................................... 26
2.2.2 Study Eligibility ............................................. 26
2.2.3 Search Strategy .............................................. 27
2.2.4 Data Analysis ................................................ 28
2.2.5 Categorisation of Simulation Training Techniques and Modalities .......................... 28
2.3 Results .......................................................... 29
  2.3.1 Search Results .............................................. 29
  2.3.2 General Results ............................................. 29
  2.3.3 Training by Surgical Modalities ............................ 32
  2.3.4 Teaching Techniques ......................................... 37
  2.3.5 Training Time ................................................ 38
  2.3.6 Numbers of Participants .................................... 39
  2.3.7 Proficiency Based Training ................................. 39
  2.3.8 Performance Assessment .................................... 40
  2.3.9 Training Outcomes .......................................... 40
  2.3.10 Quality Analysis ............................................ 41
2.4 Discussion ...................................................... 43
  2.4.1 Summary ...................................................... 46

3 Competency Based Training in Robotic Surgery: Benchmark

Scores for Virtual Reality Robotic Simulation 48
  3.1 Introduction .................................................. 48
  3.2 Methods ....................................................... 50
    3.2.1 Statistical Analysis ..................................... 56
  3.3 Results ....................................................... 56
    3.3.1 Demographics ............................................. 56
    3.3.2 Identification of Assessment Tasks ........................ 57
3.3.3 Development of Benchmark Criteria .................................. 60
3.3.4 Comparative Benchmark Assessment of Novice and Intermediate Participants ................................................. 61

3.4 Discussion .................................................................................. 63
3.4.1 Summary .................................................................................. 68

4 Evaluation of Procedural Virtual Reality Simulation Training for Robotic Surgery ................................................. 69

4.1 Introduction .................................................................................. 69
4.1.1 Objectives ................................................................................ 72

4.2 Methods ......................................................................................... 72
4.2.1 Pilot Study ................................................................................ 73
4.2.2 Participants Selection and Initial VR Simulation Training ................................................................. 73
4.2.3 Validation of the Robot-Assisted Radical Prostatectomy Modules ................................................................. 76
4.2.4 Randomised Controlled Trial of Procedural vs Basic Simulation Training ................................................................. 77
4.2.5 Power Calculation ....................................................................... 81
4.2.6 Statistical analysis ....................................................................... 81

4.3 Results .......................................................................................... 82
4.3.1 Pilot Study Results ..................................................................... 82
4.3.2 Validation of the Robot-Assisted Radical Prostatectomy Modules ........................................................................... 85
4.3.3 Skills Transfer Assessment on Fresh Frozen Cadavers ................................................................. 100
4.3.4 Skills Transfer Assessment on Dry Lab Training Models ................................................................. 101

4.4 Discussion ....................................................................................... 103
4.4.1 Summary ..................................................................................... 108
# Table of contents

## 5 A Comparative Analysis of Virtual Reality and Dry-Lab Robotic Simulation Training

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>110</td>
</tr>
<tr>
<td>5.2 Methods</td>
<td>111</td>
</tr>
<tr>
<td>5.2.1 Simulation Training Tools</td>
<td>112</td>
</tr>
<tr>
<td>5.2.2 Simulation Training Protocol</td>
<td>113</td>
</tr>
<tr>
<td>5.2.3 Power Calculation</td>
<td>113</td>
</tr>
<tr>
<td>5.2.4 Statistical Analysis</td>
<td>115</td>
</tr>
<tr>
<td>5.3 Results</td>
<td>115</td>
</tr>
<tr>
<td>5.3.1 Sensitivity Analysis</td>
<td>118</td>
</tr>
<tr>
<td>5.4 Discussion</td>
<td>120</td>
</tr>
<tr>
<td>5.4.1 Summary</td>
<td>123</td>
</tr>
</tbody>
</table>

## 6 Cadaveric Wet lab Simulation Training

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>124</td>
</tr>
<tr>
<td>6.2 Methods</td>
<td>126</td>
</tr>
<tr>
<td>6.2.1 Curriculum</td>
<td>126</td>
</tr>
<tr>
<td>6.2.2 Hands on Cadaveric Simulation Training</td>
<td>128</td>
</tr>
<tr>
<td>6.2.3 Evaluation of the Training Course</td>
<td>129</td>
</tr>
<tr>
<td>6.2.4 Statistical Analysis</td>
<td>129</td>
</tr>
<tr>
<td>6.3 Results</td>
<td>130</td>
</tr>
<tr>
<td>6.3.1 Baseline Participant Characteristics and Technical Ability</td>
<td>130</td>
</tr>
<tr>
<td>6.3.2 Assessment of Technical Performance</td>
<td>133</td>
</tr>
<tr>
<td>6.3.3 Post Course Evaluation</td>
<td>136</td>
</tr>
<tr>
<td>6.4 Discussion</td>
<td>138</td>
</tr>
<tr>
<td>6.4.1 Summary</td>
<td>141</td>
</tr>
</tbody>
</table>

## 7 Wet Lab Training: Live Animal

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Introduction</td>
<td>142</td>
</tr>
</tbody>
</table>
7.2 Methods ................................................................. 145
    7.2.1 The Live Porcine Robotic Simulation Course ............... 145
    7.2.2 Development of the Participant Survey ..................... 147
    7.2.3 Statistical Analysis ........................................ 147
7.3 Results ............................................................ 147
    7.3.1 Outcomes of Porcine Robotic Simulation Training ........ 154
7.4 Discussion ........................................................ 159
    7.4.1 Summary ....................................................... 161

8 Development and Validation of a Tool for Non-Technical Skills
   Evaluation in Robotic Surgery- The ICARS System 163
     8.1 Introduction .................................................. 163
     8.2 Materials and Methods ...................................... 165
         8.2.1 Development of the Interpersonal and Cognitive Assessment
              for Robotic Surgery (ICARS) Behavioural Rating System .. 165
         8.2.2 Validation Protocol ...................................... 167
         8.2.3 Statistical Analysis ...................................... 169
     8.3 Results ....................................................... 169
         8.3.1 Development of ICARS ................................... 169
         8.3.2 Validation of ICARS Agreement ......................... 172
     8.4 Discussion ................................................... 179
         8.4.1 Summary .................................................... 183

9 A Randomised Controlled Trial of Motor Imagery Training for
   Technical and Non-Technical Skills in Robotic Surgery 185
     9.1 Introduction .................................................. 185
     9.2 Methods ...................................................... 187
         9.2.1 Trial Protocol ............................................ 187
         9.2.2 Power Calculation ....................................... 191
9.2.3 Statistical Analysis ........................................ 191
9.3 Results .................................................... 191
  9.3.1 Technical Skill Assessment .............................. 194
  9.3.2 Non-Technical Skills Assessment ....................... 196
  9.3.3 Motor Imagery Assessment ............................... 197
9.4 Discussion ................................................ 199
  9.4.1 Summary ............................................... 201

10 Discussion ................................................. 203
  10.1 An Overview of the MARS Project ....................... 204
    10.1.1 Contribution of the MARS Project to the Science of Simulation Training ........................................ 204
    10.1.2 Systematic Review of Simulation Based Training Programmes ...................................................... 205
    10.1.3 Virtual Reality Simulation Training .................... 208
    10.1.4 Procedural vs Basic Simulation Training ................ 210
    10.1.5 Comparative Analysis of Dry Lab & VR Simulation Training ...................................................... 212
    10.1.6 Human Cadaveric Robotic Simulation Training .......... 213
    10.1.7 Live Porcine Training for Robotic Surgery ............. 215
    10.1.8 Non-Technical Skills Training for Robotic Surgery ... 216
    10.1.9 Motor Imagery Training for Robotic Surgery .......... 217
  10.2 Recommendations on Technical Skills Simulation Training for Robotic Surgery ..................................... 218
  10.3 Limitations and Future Work ................................ 222

References ...................................................... 223

Appendix A Systematic Review Search Protocol .................. 245

Appendix B A Comparative Analysis of Virtual Reality and Dry-Lab Robotic Simulation Training: Rater Training and Validation 246
# Table of contents

**Appendix C  Cadaveric Wet lab Simulation Training: Supplementary Information**  
C.1 Further Statistical Analysis .......................... 248  
C.2 Example Programme for the Cadaveric Simulation Training Course for Robotic Surgery .................................. 249

**Appendix D  Live Porcine Robotic Training Course Participant Survey Questionnaire**  

**Appendix E  ICARS Development Supplementary Data**  

248  
249  
258  
262
## List of figures

1.1 Institutions Contributing to the MARS Project ............................................. 3  
1.2 The Da Vinci Si Surgical Robot (Intuitive Surgical) ..................................... 9  
1.3 The dV-Trainer (Mimic Technologies) .......................................................... 13  
1.4 The RobotiX Mentor (Simbionix) ................................................................. 13  
1.5 Bristol TURP/TURBT Trainer (Limbs & Things, UK) .................................... 15  
2.1 PRISMA flow diagram .................................................................................. 30  
2.2 Simulation Studies Published Per Year ......................................................... 31  
2.3 Training Modalities Across Surgical Techniques .......................................... 34  
2.4 Course Participants by Surgical Modality .................................................... 35  
2.5 Training Time by Surgical Technique .......................................................... 38  
2.6 MERSQI Analysis ....................................................................................... 42  
3.1 Comparison of Participant Performance with Benchmark Scores ............... 53  
3.2 Participant Performance Analysis .................................................................. 59  
3.3 Comparison of Novice and intermediate Participant Scores to Benchmark Scores .......................................................... 62  
4.1 Example of Tubes 3 module ......................................................................... 71  
4.2 Flowchart for initial validation study ............................................................. 75  
4.3 Evaluation of Procedural Virtual Reality Simulation Training for Robotic Surgery Trial Design .......................................................... 78
4.4 Metrics Demonstrating Significant Training Effects ....... 83
4.5 Metrics Not Demonstrating Training Effects ............ 84
4.6 Participant Ratings of Realism and Anatomical Fidelity .... 88
4.7 Participant Ratings of Realism and Anatomical Fidelity .... 89
4.8 Novices Learning Curves for Bladder Neck Dissection ....... 91
4.9 Novices Learning Curves for Urethrovesical Anastomosis .... 92
4.10 Comparative performances of novice, intermediate and expert participants for Urethrovesical Anastomosis ............ 95
4.11 Comparative performances of novice, intermediate and expert participants for Bladder Neck Dissection ............ 96
4.12 Learning Curves for Basic and Procedural Training ........ 99
4.13 Comparison of Procedural or Basic VR Training vs Control ... 101
4.14 Comparison of Basic, Procedural and No VR Training ....... 102
4.15 Dry Lab Skills Transfer Assessment: Robotic Suturing Checklist Score ............................................. 103
4.16 Dry Lab Skills Transfer Assessment: GEARS Score .......... 104
5.1 Dry Lab vs VR Training Study Flow Diagram .............. 114
5.2 Comparative Performances Following Dry Lab and VR Simulation Training .................................................. 117
5.3 Comparative Performances Following Dry Lab and VR Simulation Training: Randomised Participants Only .......... 119
6.1 Change in GEARS Score Stratified by Dry Lab Experience .... 135
6.2 Change in GEARS Score Stratified by Robotic Surgical Experience 136
6.3 Opinions on the Fidelity of Cadaveric Simulation Training .... 137
7.1 Example of Porcine Robotic Surgical Training ................ 146
7.2 Aalborg Participants’ Training Experience .................. 148
7.3 Pre-Course Self Rated Robotic Skill .................................. 150
7.4 Pre-Course Robotic Console Experience .............................. 152
7.5 Pre-course Training Experience ........................................ 153
7.6 Participant’s Perceived Usefulness of Procedural Training and Over-
all Realism of Porcine Training ............................................ 155
7.7 Change in Participants’ Self Rated Ability and Confidence ........ 156
7.8 Post Course Training and Robotic Experience ....................... 158
8.1 The ICARS Rating System ............................................... 171
8.2 Expert Assessment of Feasibility, Acceptability and Educational
Impact of the ICARS Rating System ....................................... 173
8.3 Bland-Altman Analysis Comparing ICARS with NOTSS .......... 175
8.4 Comparison of ICARS Between Novice, Intermediate and Expert
Participants ........................................................................ 176
9.1 Motor Imagery Randomised Controlled Trial Study Protocol ..... 188
9.2 Motor Imagery Randomised Controlled Trial Consort Flow Diagram193
9.3 Motor Imagery Randomised Controlled Trial Technical Performance
of MI Training and Control Groups .......................................... 195
10.1 The MARS Project Robotic Surgical Simulation Curriculum ..... 219
C.1 Histogram of Total GEARS Score ..................................... 248
C.2 Normal P-P Plot ............................................................. 249
C.3 Linear mixed model assessing pre-post GEARS score controlling
for dry lab simulation and robotic surgical experience ............... 250
C.4 Example Programme for the Cadaveric Simulation Training Course
for Robotic Surgery ............................................................. 251
C.5 Example Programme (cont’d) ............................................. 252
C.6 Example Programme (cont’d) ............................................. 253
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.7</td>
<td>Example Programme (cont’d)</td>
<td>254</td>
</tr>
<tr>
<td>C.8</td>
<td>Example Programme (cont’d)</td>
<td>255</td>
</tr>
<tr>
<td>C.9</td>
<td>Example Programme (cont’d)</td>
<td>256</td>
</tr>
<tr>
<td>C.10</td>
<td>Example Programme (cont’d)</td>
<td>257</td>
</tr>
<tr>
<td>D.1</td>
<td>Aalborg Participant Survey Questionnaire</td>
<td>259</td>
</tr>
<tr>
<td>D.2</td>
<td>Aalborg Participant Survey Questionnaire (cont’d)</td>
<td>260</td>
</tr>
<tr>
<td>D.3</td>
<td>Aalborg Participant Survey Questionnaire (cont’d)</td>
<td>261</td>
</tr>
<tr>
<td>E.1</td>
<td>Comparison of Novice, Intermediate and Expert ICARS Performance Scores</td>
<td>262</td>
</tr>
</tbody>
</table>
## List of Tables

1.1 Translational Outcomes of Simulation-Based Learning . . . . . . . . . 19

2.1 Surgical Modalities Included in Training Programmes . . . . . . . . 32

3.1 Comparative Competency Rates for Benchmark Standards of 90%,
75% and 60% for Time to Complete . . . . . . . . . . . . . . . . . . . . 55

3.2 Demographic Details of HOT Course Participants . . . . . . . . . . 57

4.1 Common Key Performance Metrics Showing Significant Learning
Effects . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85

4.2 Participant Demographic Details . . . . . . . . . . . . . . . . . . 86

4.3 MANOVA for Bladder Neck Dissection Task . . . . . . . . . . . . . 93

4.4 MANOVA for Urethrovesical Anastomosis Task . . . . . . . . . . . 93

4.5 Linear Mixed Model of Novice Training on Rail Road, Ring Tower
Transfer, BND and UVA . . . . . . . . . . . . . . . . . . . . . . . . . . . 97

4.6 Test - Retest Correlation Analysis . . . . . . . . . . . . . . . . . . . . . . 100

5.1 Participant Demographics and Experience . . . . . . . . . . . . . 116

5.2 Technical Performance Scores following Dry Lab and VR Training 118

6.1 Demographic Details and Surgical Experience . . . . . . . . . . . . 131

6.1 Demographic Details and Surgical Experience . . . . . . . . . . . . 132

6.2 Change in Mean GEARS Score Stratified by Dry Lab and Robotic
Surgical Experience . . . . . . . . . . . . . . . . . . . . . . . . . . . . 135
8.1 Demographic Data and Surgical Experience of Validation Study Participants .................................................. 174
8.2 ICARS Inter Item Correlation ................................................. 177
8.2 ICARS Inter Item Correlation ................................................. 178
8.2 ICARS Inter Item Correlation ................................................. 179
9.1 Participant Demographic Details, Clinical Experience and Baseline Skill ......................................................... 194
9.2 Comparative Analysis of GEARS Subscale Score for MI and Control Groups ................................................. 196
9.3 Comparative analysis of NOTSS subscale score for the MI and Control Groups ................................................. 197
9.4 Summary Results from Movement Imagery Questionnaire .... 198
10.1 Results from the MARS Project Presented to Academic Meetings 206
10.2 Results from the MARS Project Published in Academic Journals . 207
### Chapter 1

**Simulation Training in Surgical Education**

#### 1.1 Overview and Structure of the Thesis

The structure and delivery of surgical training has undergone significant changes in recent years and continues to evolve. This change has been triggered by numerous factors both external such as economic, social and regulatory changes as well as developments in the standards and constitution of medical education. Training programmes in surgery specifically have started to adapt to these new demands, partly also in response to the growing body of evidence supporting the use of simulation training. There has been greater acceptance by the medical community of modern educational practices and the recognition that experience alone does not lead to expertise[70]. This has resulted in the expansion of competency based medical training which has been adopted by medical boards around the world. Across the majority of surgical specialities, rationalisation of the existing teaching frameworks with the paradigms of competency and simulation based training have resulted in gradual but major adaptions to existing training programmes. The relatively recent development and increasing implementation of robotic surgery offers a unique opportunity. The need for specific robotic surgical training offers the possibility to embed evidence based practices for training and assessment into
the newly established education programmes. Throughout medicine, clinicians are recognising the important of establishing standards for practice to maintain patient safety. Increasingly training programmes are directed towards meeting minimum criteria to allow practice[137, 224]. The impact of learning curves on patient safety are also increasingly viewed as unacceptable further driving the need for standardised and safe training programmes. Given the limitations on exposure to robotic surgical training, this community of surgeons have embraced the need for a global standard for training[6].

This thesis represents a body of work exploring the role of simulation to support the development of robotic surgical training. The work was undertaken in collaboration with the European Association of Urology (EAU) and partners across Europe. Together this project was named The Multi-Institutional Validation and Assessment of Training Modalities in Robotic Surgery (MARS) Project. The thesis comprises a portfolio of papers that explore all key aspects of simulation training for robotic surgery. This work made original contributions in a number of key areas:

- A review of current simulation practice
- Evaluation of the role of procedural VR, live porcine simulation and human cadaveric simulation
- Development of benchmark scores for competency in VR
- Design of a NTS rating tool specifically for robotic surgery
- Evaluation of the role of MI training for robotic surgery

The individual studies which constitute the MARS project were undertaken in collaboration with a variety of institutions and robotic experts from across Europe as shown in Figure 1.1, p. 3. International collaboration was essential for the success of the MARS project, ensuring that the conclusions from this project
1.1 Overview and Structure of the Thesis

Fig. 1.1 Institutions Contributing to the MARS Project
could be extrapolated to address the growing demand for structured, high quality robotic training across the wider world. A supranational project design was also required to provide access to the necessary resources to evaluate all aspects of robotic simulation training. As a basis for this thesis, a literature review was undertaken to assess the current state of simulation as reported in the literature. The body of this thesis consists of assessments of the key components of simulation training relating to robotic surgery. Qualitative and quantitative methods were used to address each component individually and aim to build upon existing literature. Chapters 3, 4, 5 evaluate the potential of virtual reality simulation (VR) and dry lab training. Chapters 6 and 7 represent the first analysis of the role of wet lab training for robotic surgery. Non technical skills (NTS) for robotic surgery were considered in chapter 8 leading to the development of the first rating system specifically for robotic surgery. Finally in chapter 9 I report the first randomised controlled trial evaluating the role of mental training for robotic surgery.

1.2 A Brief History of Simulation Training in Surgery

Apprenticeship training has historically played a central role in western medical education. Up to the 19th century, medicine remained relatively clearly divided between the academic physicians and other more technically trained practitioners such as surgeons, apothecaries and tooth extractors. Physicians were trained in medicine (or physic) in the new universities that had arisen from the monasteries and traditional seats of learning across Europe such as Padua, Leiden, Oxford and Cambridge. A theoretical curriculum based on the works of Galen was taught with limited practical applications. Anatomical dissections, when performed, would only be used to demonstrate Galen’s often incorrect teachings rather than
for any greater scientific benefit. In contrast other medical practitioners such as those mentioned above would learn their trade through an apprenticeship model. Apprentices would often pay to be indentured to a master craftsman. Training took several years during which time the apprentices would be expected to undertake various menial tasks and errands as required by their master. Training was determined by the guilds and consequently both study time and the specific training requirements varied considerably. Over time there was an expansion in the regulation of medical training with government playing an increasingly important role in licensure of the profession trade. In 1518 Thomas Linacre established the Royal College of Physicians together with medical chairs at Oxford and Cambridge.

The early 19th century saw a major change in medical education. Students would initially undertake a course of lectures or a formal medical degree depending on whether they wished to become a physician, surgeon or apothecary. This training remained largely theoretical. Practical training was undertaken in subsequent clerkships or apprenticeships which were increasingly completed in hospitals. By 1830, clinical training could be undertaken at one of seven teaching hospitals in London or in Edinburgh, Aberdeen or Dublin[76]. For a fee, students would become apprentices to physicians, apothecaries or surgeons in the teaching hospitals often following a period of training with provincial practitioners. For those who could afford it, it was also possible to become a dresser or “cub” to a leading hospital surgeon or physician. They would assist in surgery, see new patients and be on call. Training could vary considerable according to the inclinations of the master. Over the course of the 19th century, medical schools increasingly took on formal roles of training “doctors” both pre-clinically and clinically. However the lack of a standard curriculum and disparities in training received resulted in the formation of the General Medical Council, originally the General Council of Medical Education and Registration, in 1858. Major concerns were the prioritisation of
theoretical over practical training and wide variations in admission and licensing bodies[267]. The resultant report of the GMC’s Education Committee set out a curriculum for medical training lasting 5 years with specific requirements for chemistry, physics and biology but little increase in clinical training.

In contrast a significantly more refined and modern system was being practised in Germany and across the wider western world. Originally developed in the 18th century by Herman Boerhaave, Professor of Medicine at the University of Leiden, the German system became widely recognised as at the forefront of medical education. It consisted of a close integration between the basic sciences and clinical medical training alongside a relatively structured clerkship programme. Teaching was coordinated by full time academics and intense competition was fostered amongst trainees with only the best and most dedicated progressing to a position working with the professor[210]. Most famously this system of training was developed further and introduced to John Hopkins Medical School by William Halsted. Halstedian training, which has now become synonymous with the apprenticeship model of medical education, focussed on the graduated responsibility given to trainees as they advanced in addition to intense and repetitive opportunities for treating patients and an understanding of the scientific basis of disease. Less well known is the intense competition fostered by Halsted with only very few of the best trainees progressing to become residents and his uncompromising approach to standards[182]. Trainees had to be available any time of day or night 365 days per year and there was no set length to training with Halsted deciding when a trainee was ready to practice[24].

The Halstedian system of a structured residency programme continued to be used for over a century. Whilst being criticised for the long, onerous hours especially in surgery, it remained an effective approach for training competent clinicians. Towards the end of the 20th century various factors meant that this training model was increasingly questioned. Around the world, overly long
working hours even in medicine were deemed unacceptable both for the health of the workers as well as concerns over errors and safety. Changes were made in the allowed working hours most significantly in Europe with the introduction of the European Working Time Directive that limited all workers to 48hrs per week with further controls on rest periods. Similarly, working hours were reduced in the USA with guidelines limiting resident to 80 hours per week. Another major influencing factor on surgical training were the increasing concerns over medical errors and complications. Expectations for zero-complication surgery have led to the expansion of safeguards, standardisation of practices and ever-greater scrutiny of surgical outcomes. Publication of the report “To Err is Human” which highlighted that 10% of hospital patients suffered a complication led to increasing evaluation of clinical training[253]. This issue has been highlighted in the UK by the publication of surgical outcomes for a number of specialities. As a result, the effect of learning curves on surgical outcomes, specifically with regards to trainees, has been carefully scrutinised.

A major influencing factor is the progressive pressures on healthcare budgets. Rising demand and increasing healthcare costs are putting increasing pressure on healthcare budgets. As a result, greater efficiencies throughout the industry including both training and operating room efficiency.

Despite the extensive technical skills, NTS and underlying knowledge necessary for surgery, until recently, little was known or studied about the process of surgical skill acquisition. Out of the Halstedian model of surgical apprenticeship, a three-stage process was broadly adopted for surgical skill training. Initially trainees would just observe a number of surgical procedures. In the second stage they would perform the techniques under close supervision. Finally in the third stage they would undertake a more independent role as the main surgeon[57]. Whilst not an accurate description, this process is widely known by the phrase “see 1, do 1, teach 1”. However, it was increasingly recognised that for safe clinical practice and
efficient surgical training, surgical training needed to be performed in a dedicated
environment and often outside the operating room away from “real” patients.
Another important development has been the realisation of the importance of
focused training. Achieving aptitude in everyday tasks to an acceptable level
such as learning to drive or play recreational golf is relatively easy to achieve with
limited training and practice. It has been estimated to take less than 50 hours for
most skills[69]. At this stage, an automated state is reached in which the task
can be executed relatively smoothly with infrequent errors. In contrast it is now
recognised that development of expertise rather than just aptitude in a particular
skill or field is not solely a product of the length of experience or training. It
has been shown that it takes at least 10 years training to reach the highest level
in a discipline[70]. Furthermore continued engagement in deliberate practice is
necessary to ensure that skills continue to improve.

Simulation training is the key to addressing these questions. Both for the
novice and expert, the role of simulation training is increasingly being recognised.
In the initial stages of the learning curve, simulation enables surgeons to gain
the relevant experience before encountering their first live patient. Further in
the learning curve, simulators provide the opportunity for deliberate practice,
focussing practice on areas of weakness. Like an international violinist who will
spend countless hours practicing difficult pieces, the surgeon is able to hone his
skill outside the operating room.

1.3 Robot Assisted Laparoscopic Surgery

Robot assisted laparoscopic surgery (hitherto referred to as robotic surgery)
entered general surgical practice following the US Food and Drug Administration
(FDA) approval of robotically assisted surgical devices for human surgery in 2000.
At the time of writing, the Da Vinci surgical robot (Intuitive Surgical, Sunnyvale,
1.4 Simulation Training in Surgery

CA, USA) remains the only surgical robot licensed for use on human beings both in Europe and the US. The Da Vinci surgical robot offers the surgeon a number of unique benefits such as 3D vision, 7 degrees of freedom of laparoscopic instruments, tremor damping, motion amplification and camera stability. It is argued such advantages make the Da Vinci surgical robot superior to open and laparoscopic approaches particularly in delicate surgery such as radical prostatectomy. These claims are yet to be rigorously defended but this has not limited the expansion with over 4,400 robots in use globally in 2017. The major drawback of robotic surgery remains the high capital costs and the price of the ongoing service contracts and consumables.

Fig. 1.2 The Da Vinci Si Surgical Robot (Intuitive Surgical)

1.4 Simulation Training in Surgery

In the past two decades there has been an extensive expansion in the both the availability and application of simulation training models in surgery. To a large extent, these changes have been driven by developments in other fields such as the aviation industry. A wide range of surgical simulation modalities have been
developed and are increasingly being incorporated into the surgical curriculum. Simulators have been developed for both technical and non-technical skills (NTS) in surgery.

1.4.1 Virtual Reality Simulation Training

VR simulation was introduced relatively late to surgical training in comparison to the aviation industry. VR was first applied to surgical training with the development of a general abdominal simulator by Satava [212] in 1991. VR simulation remains one of the most technologically advanced methods for surgical training. It has been enthusiastically embraced by units around the world but for a long period of time the medical community resisted its integration. Numerous high-quality studies demonstrating its effectiveness have helped to overcome this scepticism. Landmark studies have shown that VR training results in significantly better operating room performances for laparoscopic cholecystectomy [219, 87]. VR simulation training has successfully been applied to a wide variety of endoscopic surgery such as gastrointestinal endoscopy [122], neurosurgery [124], endovascular surgery [217] and gynaecology [131, 78]. In all cases VR simulation was found to be at least as effective as other training modalities or traditional surgical training.

Alongside the basic laparoscopic skills simulators discussed above, a number of VR simulators have been developed for urological surgery. The laparoscopic simulators Procedicus MIST (Mentice, Sweden), LAP Mentor (Simbionix, USA) and LapSim (Surgical Science, Sweden) all offer modules on laparoscopic nephrectomy. Consequences of training have only been demonstrated with the LapSim with successful transfer of skills to the OR [11]. Interestingly a study by Wijn et al. [261] demonstrated that the Procedicus MIST lacked construct validity and was not an effective training tool: a rare example of a negative trial in surgical education.
Alongside laparoscopy, a variety of VR simulators are available for other modalities including cystoscopy, ureteroscopy, percutaneous nephrolithotomy and transurethral bladder tumour and prostate resection. For urolithiasis training, the URO Mentor (Simbionix, USA) has been most widely tested. However, as commonly found in simulation research, the validity evidence is primarily focussed on the content and construct with expert-novice comparisons. Transfer of skills to the OR has been shown for cystoscopy[214]. Similarly, ureteroscopy training has shown transfer for skills to cadaveric performance[146, 168]. A wide variety of transurethral resection VR trainers are available. SurgicalSIM TURP has undergone the most extensive validation assessment but analyses are limited to content, construct and face validity; no assessment of the predictive validity of the training tool have been performed[202, 236]. Alongside resection trainers, further models for Greenlight, thulium vaporesection and enucleation, urolift and transurethral thermal water vapour therapy are also available.

In comparison to laparoscopic and endourological surgery, a relatively large number of VR simulators have been developed for robotic training. The principle VR simulators available for robotic surgery are the Robotic Surgical Simulator (ROSS) (Simulated Surgical Systems, USA), the dV-Trainer (Mimic Technologies, USA) (Figure 1.3, p. 13), SimSurgery Educational Platform (SEP) Robot (SimSurgery, Norway), da Vinci Skills Simulator (dVSS) (Intuitive Surgical, USA), ProMIS (CAE Healthcare, Canada) and RobotiX Mentor (Simbionix, USA) (Figure 1.4, p. 13). The majority of these have undergone extensive validation. Both simulator hardware and software vary considerably between the different models. The dVSS is the only simulator to work directly with the da Vinci robot. The dVSS backpack is attached directly onto the console, enabling the user to practice operating on the da Vinci robot in a virtual environment. All the others are standalone simulators. Whilst mimicking the Da Vinci robot, the standalone dV-Trainer hand controls differ from those of the da Vinci system, with the
master controllers connected via two tension cables as opposed to the jointed arms of the da Vinci robot. Likewise, the RobotiX Mentor, released in 2014, uses free-floating hand controls. The SEP robot uses two motion-tracked hand controls that mimic rather than replicate robotic control arms. Like the da Vinci robot, a clutch is incorporated, but the video feed is displayed by a 2D screen as opposed to the 3D video provided by the above simulators. Regarding assessments of validity, the majority of validity evidence relates to content\cite{17}. Relationships to other variables was demonstrated in a number of studies. Moderate relationships between simulator scores and related tasks on the surgical robot where shown in six studies\cite{106, 135, 178, 65, 201, 107}. Strength of association ranged between $r=0.55 – 0.87$. Importantly a number of studies have reported evidence to support the consequences of VR robotic training. Notably Culligan et al.\cite{55} demonstrated that completing a training programme using the dVSS simulator led to successful completion of a supravesical hysterectomy equivalent to experienced robotic surgeons measured by blood loss and operative time. Global assessment tool for evaluation of intraoperative laparoscopic skills (GOALS) used to assess technical proficiency was equal between novice surgeons trained on VR and establisher robotic surgeons.

To date the majority of VR simulators that have been assessed have provided basic robotic skills training and suturing practice. Modules use abstract exercises such as placing hoops on pegs or cutting a shape out of a material to train specific skills. These include both generic (e.g. hand eye coordination) and robot specific skills for example clutching. More recently procedure specific VR training modules have been developed for a number of simulators. Of the six VR simulators available for robotic surgery, procedural training is available for three (RoSS, the dV-Trainer, the RobotiX Mentor). Prostatectomy, cystectomy, and lymph node dissection training is available for the RoSS simulator. Only basic content validation has been undertaken\cite{48}. Mimic have developed an alternative system
1.4 Simulation Training in Surgery

Fig. 1.3 The dV-Trainer (Mimic Technologies)

Fig. 1.4 The RobotiX Mentor (Simbionix)
called Maestro which provides procedure-specific training through manipulation of a 3D anatomical video[107]. Content validation has been undertaken but the programme does not allow full procedural training. The most advanced system is available on the RobotiX mentor. This programme currently allows training on prostatectomy and hysterectomy. It remains unvalidated but offers the potential for bridging the gap between basic simulation training and advanced wet lab and modular training within the operating room.

1.4.2 Dry Lab Simulation Training

Dry lab models are widely used in surgical simulation particularly for open and laparoscopic training. In laparoscopic surgery box trainers (dry-lab simulation models) have been shown to be largely equivalent to VR simulation[163]. They allow training using actual laparoscopic instruments and provide realistic haptic feedback. However they lack the objective performance assessment provided by VR simulation. The fundamentals of laparoscopic surgery (FLS) program is a widely validated basic skills training programme developed for box trainers. Whilst the consequences of FLS training is yet to be established, a number of validations studies have been published supporting content, internal structure, relationships with other variables and response processes[272]. The major advantages of dry lab models are that they offer relatively low cost, effective training especially for basic skills that is easily accessible[241].

Specifically in urology, dry lab training models are used extensively in endoscopic training. Models are available for TURBT/TURP (Bristol TURP/TURBT Trainer (Limbs & Things, UK) (Figure 1.5, p. 15; Resection-Trainer (Samed GmbH, Dresden, Germany)), ureteroscopy (Cook URS model (Cook Medical, USA); K-Box (Porgés-Coloplast, France); Uro-Scopic Trainer (Limbs & Things, UK)) and percutaneous nephrolithotomy (the Perc Trainer (Mediskills); PCNL
1.4 Simulation Training in Surgery

Trainer (Limbs & Things)). Most have only undergone content validation although relationships with other variables has been shown for the Uro-scopic trainer with training resulting in significant improvements in performance[30].

For robotic surgery, a number of dry-lab models have been developed for use with the da Vinci Surgical System but only two basic skills models have been formally validated. Anecdotally most dry lab training is undertaken using basic skills models such as suture pads. A set of three dry-lab models were reverse engineered from the Mimic Msim VR software[201]. Similarly Goh et al. [84] developed four training exercises (suturing, dissection, peg transfer, and needle driving) for robot-assisted radical prostatectomy (RARP). Validation in both cases extended only to expert novice comparisons. Several procedure specific models have also been described in the literature. The SIMPLE partial nephrectomy

Fig. 1.5 Bristol TURP/TURBT Trainer (Limbs & Things, UK)
model was developed using 3D printing comparisons between performances on the model and in live surgery were shown to be equivalent demonstrating relationships with other variables[41].

1.4.3 Wet Lab Simulation Training

Both live animals and cadaveric human tissues provide unique opportunities for training. Anatomical fidelity and the ability to train complex procedure specific skills together with complication management mean that they continue to form an important component of surgical training. Aside from ethical concerns, further disadvantages are the extensive facilities and costs necessary to provide such training. Ex-vivo and live animal models have undergone limited assessment in the literature to date. Live porcine models have been used for ureteroscopy training albeit without meaningful validity analysis. Similarly, human cadavers have been utilised for ureteroscopy, laser prostatectomy and cystoscopy training[29, 6, 172]. Likewise laparoscopic radical prostatectomy has been undertaken using live porcine models[209].

For robotic surgery, whilst live animal training is regularly provided by numerous centres across Europe and the US, training using either live animals or cadaver has not been formally assessed. Live porcine models form part of the European Association of Urology (EAU) robotic training curriculum. Although specific data pertaining to their use as wet lab models are not available, the curriculum was shown to be effective overall good educational impact and acceptability[254].

1.4.4 Motor Imagery Training

Cognitive training in the form of mental or motor imagery has been explored as a training adjunct in a variety of fields. Motor imagery (MI) is defined as the mental rehearsal of actions without performing in the movements involved[159]. The
utility of MI has been seen in various fields from sport and music to rehabilitation medicine. A number of studies have demonstrated the utility of MI training for surgical training[255, 58]. Results generally are positive but the majority of studies suffer from poor methodology limiting the conclusions that can be drawn. Outside of surgery, there is a large literature base supporting MI. Patterns of neural activation during MI have been shown to closely correlate with the reciprocal motor tasks and MI has been found to stimulate brain plasticity and learning[88]. Furthermore, studies in sports and cognitive psychology have led to evidence based models for effective MI training in a number of fields[159].

Yet despite extensive research, the factors that enable successful implementation of MI remain debated[64]. As mentioned above many studies are limited by poor methodologies and results do not appear reproducible. Furthermore, no studies have been undertaken for complex surgery such as robot assisted laparoscopy. MI offers many potential benefits to surgical training. It allows low cost training and rehearsal outside the operating room whenever it is convenient for the trainee. To date the majority of training has focussed on technical skills but the ability of MI to train non-technical skills also needs to be explored.

1.4.5 Non-Technical Skills Training

It has become well established that the technical performance during surgery accounts for little of the overall risk of error and complications. It has been shown that whilst 50-67% of errors are preventable, cognitive factors such as errors in judgement or vigilance contribute to 86% of mistakes and other non technical factors such as communication and Interruptions/distractions are involved in 83%. In comparison inexperience and a lack of competence was only associated with 53% of errors[79].
Compared to other industries, medicine has lagged far behind in this important field. In the 1970’s, NASA developed a behavioural training system, known as Crew Resource Management (CRM) Training. Its objective was to enhance positive behavioural domains whilst reducing the likelihood of error occurrence, and it later became standard training in aviation. In addition the persistent misconception exists that nontechnical skills are inherent behaviours rather than skills that require specific training, practice and assessment has also only recently been overturned[270]. A number of terms are used for NTS. There is ongoing debate as to whether “non technical” is a misleading term but it continues to be widely used in the surgical literature and will be the term used in this thesis[165].

NTS can broadly be divided into three categories: social (communication, leadership, teamwork); cognitive (decision making, planning, situation awareness); personal resources (ability to cope with stress/fatigue)[269]. Training of NTS may incorporate a number of different techniques such as lectures or demonstrations but simulated exercises are most commonly used. Whilst simulated performance of the task are important like in technical simulation, feedback and discussion of the performance forms a central component of the training process. Rating systems are important tools allowing structured feedback as well as assessment of trainees. A wide variety of rating systems have been for assessment of the entire surgical team as well various individual team members such as surgeons, anaesthetists and scrub practitioners[264]. For surgeons, the gold standard training tool is NOTSS[269]. However to date no specific training tools have been developed for robotic surgery.

1.4.6 The Effectiveness of Surgical Simulation Training

As a result of the large volumes of research on simulation, there is a very broad body of evidence available to evaluate the effectiveness of surgical simulation. As
Table 1.1 Translational Outcomes of Simulation-Based Learning

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal acceptability</td>
<td>Participant satisfaction with the simulator/assessment</td>
<td>Post-training questionnaire or performance self assessment</td>
</tr>
<tr>
<td>Contained effects</td>
<td>Effect of the training or other educational intervention on performance</td>
<td>Global rating scale for technical skills assessment</td>
</tr>
<tr>
<td>Downstream effects</td>
<td>Changes in clinical behaviours</td>
<td>Transfer of knowledge/skills to clinical practice</td>
</tr>
<tr>
<td>Target effects</td>
<td>Effects of simulation/educational intervention on patient outcomes</td>
<td>Reduced complications or improved post operative outcomes</td>
</tr>
</tbody>
</table>

discussed in section 1.4.7, the form as well as the content of this evidence is very important. The outcomes of simulation training may be assessed on a number of levels. Immediate effects on the participants, both subjective (opinions on training and perceived effects of training) and objective (for example technical or non technical skill assessment) are widely reported. The aim of simulation may be considered to be to replace or support real experiences in order to support training. In medicine therefore, whilst direct effects are important, so are the more distant translational outcomes; better patient outcomes, safer patient care, cost saving [149]. The potential outcomes of simulation training are shown in Table 1.1, p. 19.

A core assumption of surgical simulation is that training results in improvements in skills that are directly transferable to the OR[234]. To address this question, a number of high quality systematic reviews have analysed the available evidence. Unsurprisingly there is consistent evidence to support the use of simulation over no training although most data related to contained effects[234, 59, 53].
Significantly fewer studies evaluated downstream effects of surgical simulation, but pool analysis have shown that simulation training may result in significant improvements in OR performance\[53, 234, 59\]. These reviews did also note the increased heterogeneity in the results of such skill transfer analyses. This in part reflects the variations in the training paradigms applied across the studies. The lack of clear methodologies prevented clear analysis but variations in the length of training, intensity of training and the form of training (for example use of deliberate) were seen as possible factors across different reviews\[59, 234, 53, 149\]. The heterogeneity of the results have also prevented further assessments on the length and degree of training required for optimal skill transfer to the operating and questions on skill retention also remain to be answered.

### 1.4.7 Validation Assessment for Simulation Training

As outlined above, there is already a considerable breadth of literature evaluating the potential of simulation training in surgical education. High quality and comprehensive reviews, including those by Cook et al. [50, 51], Dawe et al. [59] have clearly reported on the current available evidence and its limitations. One critical and pervasive area of weakness is that of the validation for simulation training. The process of assessing and validating educational interventions is crucial especially in the medicine. Increasingly policy makers, regulators and the general public are demanding the assessment and maintenance of clinical performance. This requires both effective training as well as accurate assessment\[52\]. Validation is a critical component in this evaluation of education interventions and assessments. It refers to the systematic process of collecting evidence to support or oppose the use of an educational instrument or test. The majority of the surgical literature focuses on the validation of simulators. The past four decades have seen advancements in the validity theory. The first consensus statements on the
process of validity were published by various bodies as early as 1954\cite{18}. The classical framework of validation, published in 1974, was widely adopted by the medical education community. It continues to inform the evaluation of educational interventions and simulators particularly in surgery\cite{167}. The framework was characterised by distinct types of validity that could be individually assessed. Common sources of validity applied to surgical simulation, for example face (how closely a simulation/assessment covers the concepts or resembles the task it relates to), content (how well the simulation/assessment covers all aspects of the construct being assessed/simulated) and construct (the ability of the simulation tool/task to differentiate levels of ability), continue to be commonly used in the literature. Yet these recommendations on the structure of validation assessments were superseded by the unified concept of content validity published in the 1999 APA/AERA consensus standards\cite{66}. A more holistic approach is advised whereby all validity is considered construct validity with evidence from five sources used to refute or support the specific application of an educational intervention in a specific environment. Importantly this new approach to validation no longer supports the concept of determining whether a simulator is valid- validity should not be considered a fixed property and may vary according to the situation in which it is being used. Secondly individual components of validity assessments should not be considered isolation. Given the need to consider multiple sources of evidence relating to the specific implementation of a simulation tool, individual assessments such as face validity which remain ubiquitous in the literature, provide very little useful evidence to support a simulation tool or assessment. Most of the published literature relating to surgical simulation pertains to the outdated 1974 validity framework. Even then the majority of papers offer only limited validity evidence \cite{50}.

A number of assessments evaluating the validation of simulation based training have been undertaken\cite{50}. Large numbers of validation studies have been published
but the classical framework of validation, published in 1974, continues to be commonly applied to the assessment of medical education and simulation. Yet the recommendations on the structure having been superseded by the unified concept of content validity in the 1999 APA/AERA consensus standards. Yet the classical framework continues to inform the majority of validation studies. In addition a large proportion of studies offer only one source of validity evidence and 33% of studies reported only relations with a learner characteristic. In contrast less than 5% of studies provided evidence for either response process or consequences. Cook et al. [51] reported only 7% (n=28/417) studies looked at associations between simulation training and real patient outcomes.

Similarly Dawe et al investigated skills transfers following simulation training. 82% (n=23/28) demonstrated a positive effect of simulation training over no training. Four studies compared simulation to patient based training, with all showing that simulation was not inferior and two studies that utilised a comprehensive curriculum of simulation training found it to be superior[59]. Again authors note the poor methodological quality in a significant proportion of papers.

1.4.8 Implementation of Simulation into Surgical Training

Effective implementation of simulation training into the surgical curriculum relies on a number of factors. Care needs to be taken in the design of a surgical curriculum or training programme. It is important to consider the learners or participants, the learning environment and how training will be delivered. When considering delivery of training, the training modality also needs to be carefully considered (high vs low fidelity; VR vs dry lab vs wet lab). Training needs to be tailored to the experience and seniority of the trainee. Junior trainees at the beginning of their learning curve require training in basic skills. Such basic skills
training can well be undertaken using basic training models that allow the training of generic surgical skills such as hand eye coordination, needle driving, camera manipulation (in the case of endoscopic or laparoscopic surgery) etc. Training should focus on these skills to allow trainees to master these basic concepts and therefore do not necessarily require them to be procedural. For intermediate skilled trainees more procedural training is required. It has been proposed that full procedural training could be introduced at this stage[16]. For experienced trainees, simulation has been advised to centre on advanced skills such as crisis management.

Assessment requires careful consideration. The most direct method for assessing surgical training is patient derived outcomes; does training improve outcomes for patients. However, given the inherent complexity of any patient’s journey, direct comparisons between surgical training and final outcomes are very difficult. As a result, surrogate markers are used to measure technical performance with the assumption that these will then correlate to a certain degree with patient outcomes. These include both simple, direct measures such as time, scoring systems such as error-based checklist and global rating scales. For non-technical skill similar considerations need to be made. For NTS, rating scales and checklists are almost exclusively used both for performance assessment and to help guide subsequent debriefing.

The ultimate goal of surgical simulation is to improve clinical performance and outcomes. Transfer validity has been demonstrated for VR simulation in robotic surgery but this relies on competent performance in the simulation environment[219]. Determination of competency entails objective performance assessment and setting standards against which trainees can be measured. Surgical simulators provide a wide range of data on an individual’s performance in addition to the various external scoring systems that have been developed such as GEARs. Yet without set benchmarks, such results remain abstract and
unrelated to clinical outcomes. Using the scores of expert surgeons as a marker for competency simulation training to clinical benchmarks of proficient performance. Classically expert panels are used to in standard setting. Various approaches such as the Angoff or Hofstee method may be used to set the pass score through a collective expert decision on the acceptable performance. However the inherent subjectivity and potential for bias limits their applicability[117].

1.5 Aims of the MARS Project

The aim of this thesis is to identify, assess and validate the key components required for the delivery of effective robotic surgical simulation. Using a mixed methods, multi institutional approach the MARS project will address the role of basic (VR, dry lab) and advanced procedural (VR, wet lab) simulation training in addition to NTS and mental training. The ultimate aim of the MARS is to provide evidence for the development systematic training curriculum in robotic surgery for residents and specialists across Europe and the wider world.
The State of Training in Surgery: A Systematic Review

2.1 Introduction

Simulation tools have become widely and extensively integrated into medical curricula across the majority of specialities. Within surgery, simulators have been developed for a wide variety of surgical procedures and techniques utilising the breadth of simulation modalities available. Numerous, comprehensive systematic reviews and meta-analyses have further assessed the quality of this evidence as discussed previously.

Effective simulation training tools are central to the successful implementation of training programmes. Simulation is only one of a range of teaching strategies available to educators and the benefits that result from using such tools are contingent on their appropriate integration in the training programme or curriculum[160]. Surgical simulators are developed in order to support the learning of a particular task or skill. Surgeon educators must identify the education skill goal of the training and understand the requisite domains of knowledge to enable appropriate integration of simulation into the training programme[45]. At best unstructured training will not be efficient and at worst risks negative training whereby the
trainees practice bad habits[77]. Furthermore, the simulator must be suited to the training task. For example, basic simulators can be very effective for basic skills training resulting in generalizable learning however the scope for transfer from more advanced procedure simulation training is more limited.

The extensive evidence that has been accumulated in the literature in the evaluation and validation of simulators is important to support the development of effective surgical training. Alongside evaluation of the simulation tools, the design and execution of training programmes themselves must also be considered. The aim of this review is the evaluate the quality of surgical education programmes through published reports.

2.2 Methods

This systematic review was planned, conducted and reported according to the PRISMA standards of quality for reporting meta-analyses[157].

2.2.1 Aims

This systematic review aimed to describe, compare and contrast training methodology and techniques used in the implementation of simulation tools into surgical training programmes.

2.2.2 Study Eligibility

Broad inclusion criteria were used to ensure all relevant studies were identified. Studies published in English or German were included if they described a training protocol involving surgical simulation tool(s) for the teaching of any healthcare professional at any stage of practice. A surgical simulation tool was defined by the authors as any form of physical simulation or simulator used to train an operative
2.2 Methods

surgical procedure or technique. Results were limited to general surgery, urology and mixed speciality training (defined as any non-speciality specific training for generic surgical skills). Bedside procedures (not routinely occurring within an OR) such as central line insertion or needle thoracostomy were excluded. The focus of this review is the training of intraoperative surgical skills. Reports of training that included preoperative and post-operative patient management were included only if the intra-operative portion of the training met the inclusion and exclusion criteria. Training for preoperative or post operative healthcare interventions were excluded.

Trials were excluded if specific details on the training protocol were not provided or if the trials did not explicitly state training as one of the primary objectives of the study. For example, validation studies in which the performance of participants was analysed without explicit details of any training were excluded. Similarly if studies analysed the performance following training without reporting details of the training were also excluded. Data collection were performed independently by two study investigators. Any disagreements were resolved through discussion with a third study investigator. Correlation between MERSQI scores was assessed to determine agreement using Cronbach’s alpha.

2.2.3 Search Strategy

Searches of MEDLINE and EMBASE databases were conducted to identify all relevant studies. Searches were completed on 1st September 2018. The following search terms were used in various combinations (for full search strategy see appendix A). Reference lists were hand searched to identify relevant texts.

All retrieved studies were evaluated against inclusion and exclusion criteria. Data was extracted and compiled on an electronic spreadsheet. Data was collected on study characteristics (time and place of study), participants, type of training,
2.2 Methods

design and rationale for training regime, training assessment, training duration (converted to total hours with the assumption that a day of training equated to 8 hours) and outcomes.

2.2.4 Data Analysis

Qualitative analysis of the data was undertaken. Given the heterogeneity of the study designs and outcome measures, pooled analysis could not be performed. All studies were assessed using MERSQI, a validated tool to measure the quality of medical education research[204]. Studies scoring at least 14/18 were considered high quality[183].

2.2.5 Categorisation of Simulation Training Techniques and Modalities

To aid the description and analysis of the results, a number of categories were used throughout the review. In accordance with the standard practice in academic surgical education, simulators were grouped into 4 principle categories. Virtual reality (VR), dry lab (any synthetic, non-organic model), wet lab (use of live or cadaveric animal or human tissue).

Training itself was categorised as either basic or procedure specific. Basic training was defined as the training of generic surgical skills. Procedure specific training was defined as any that focuses on a specific surgery or part of a procedure. These divisions were selected given the differing requirements for simulation training models. Basic training models may bear little resemblance to the actual anatomy but should closely replicate the technical task. In contrast, procedure-specific models should offer greater fidelity to the corresponding anatomy and tissues[60].
Participants were divided according to their training stage. In accordance with European and American custom, interns were defined as junior doctors not undertaking specific surgical training. Residents were subcategorised by their training year (junior residents = post graduate year 2-3; senior resident = PGY 4+; if authors reported participants as junior/ senior residents or the equivalent this was followed). Attending or other fully trained, independently practicing surgeons were classified as consultant surgeons.

Teaching techniques were divided into four groups. Explicit knowledge-based learning such as lectures, e-learning, reading; proctored training in which a trainee is directly taught technical skills by a mentor; self-directed training in which participants practice without assistance; observation in which participants learn by observing an expert perform a technical task; debrief in which a performance is reviewed and discussed with the trainee post hoc.

2.3 Results

2.3.1 Search Results

Following initial searches 2614 studies were identified. 969 abstracts were reviewed and 137 studies were included in the final review (Figure 2.1, p. 30). Agreement in MERSQI score between the two investigators collecting data was high (Cronbach \( \alpha = 0.76 \))[129].

2.3.2 General Results

Studies were published between 1992 and 2017 with the majority (64%) published since 2012 (Figure 2.2, p. 31). Studies were drawn from a across the world. USA, UK and Canada accounted for 63% of all studies with the USA alone accounting for 42%. 
2.3 Results

PRISMA Flow Diagram

- Identification
  - 2614 Studies identified through database searches

- Screening
  - Titles Screened & 969 Studies Identified
    - Study not relevant = 458
    - Full text not available/abstract only = 9
    - Incorrect Language = 2
    - No original data = 6
  - Abstracts Screened & 494 Studies Identified
    - No specific training programme/curriculum described = 158
    - Insufficient detail to assess training = 70
    - Incorrect Speciality = 130
    - No surgical simulation described = 12

- Eligibility
  - Full Papers Screened & 124 Studies Identified
    - Added following hand searches of reference lists = 13

- Included
  - 137 Paper Included in Final Review

Fig. 2.1 PRISMA flow diagram
2.3 Results

Fig. 2.2 Simulation Studies Published Per Year
Table 2.1 Surgical Modalities Included in Training Programmes

<table>
<thead>
<tr>
<th>Modality</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laparoscopy</td>
<td>86</td>
</tr>
<tr>
<td>Robotic Surgery</td>
<td>26</td>
</tr>
<tr>
<td>Open</td>
<td>22</td>
</tr>
<tr>
<td>GI &amp; GU Endoscopy</td>
<td>14</td>
</tr>
<tr>
<td>Microsurgery</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>151</strong></td>
</tr>
</tbody>
</table>

A variety of different teaching models were employed to provide training covering all the principle surgical modalities and approaches (Table 2.1, p. 32). The majority of studies offered laparoscopic training either alone or in combination (n=86, 63%). 11 (8%) of studies reported on courses offering training in more than one surgical modality.

Training of both basic and advanced procedure-specific skills was reported with almost half the studies (n=65, 47%) focussing on just basic skills training. The remaining training programmes provided either procedural alone or both basic and procedural-based training. A variety of methods and approaches were used to provide training. Most studies (n=79, 58%) included knowledge-based training. For technical skills, 68% of programmes (n=93) used dry-lab models, 57 (42%) included VR simulation training and 36% (n=49) used animal (live or cadaveric) or cadaveric human tissue for training. 52 (38%) studies used more than modality for technical simulation training.

2.3.3 Training by Surgical Modalities

Laparoscopic Training

Laparoscopic surgery was the focus of the majority of training programmes with 86 (61%) studies reporting laparoscopic training. The majority focussed on basic
2.3 Results

laparoscopic skills training (n=48) while 19 programmes offered only procedural based training and 19 offered a combination of basic and procedural training.

In total 67 programmes offered some form of basic laparoscopic skills training (Figure 2.3, p. 34). 40 (59%) utilised a single simulation modality (VR=12, dry lab=28). 26 programmes used a combination of training modalities. 17 used dry-lab with VR and 7 dry-lab with wet-lab. 2 programmes offered dry-lab, VR and wet lab training. 1 study did not report the training modality for basic laparoscopic training[27].

38 courses offered procedure-specific training. 18 courses used only wet lab simulation with a further 9 VR only and 6 dry lab only courses. 3 courses used wet lab with dry-lab training models and 1 course used VR, dry lab and wet lab simulation. 1 course did not report the modality used for procedural training[211].

Laparoscopic training was provided to a wide variety of participants. The majority of courses provided training to inexperienced participants. 22/86 courses trained only medical students or interns. A further 36 courses also included junior residents. Only 14 courses included consultant surgeons and a further 25 included senior residents (PGY4+). 10 did not provide details on the training group. In contrast procedural training courses did include a higher portion of more experienced participants (68% (n=13) included senior residents or consultant surgeons) in comparison to basic laparoscopy courses (Figure 2.4, p. 35).
2.3 Results

Fig. 2.3 Training Modalities Across Surgical Techniques
Fig. 2.4 Course Participants by Surgical Modality
Robotic Surgery

26 courses provided basic and/or procedural based training with a further two including robotic surgery alongside laparoscopy and cystoscopy. Most courses provided only basic skills training (n=16). 9 courses offered basic and advanced simulation and 1 provided only procedural training.

VR simulation was used for the majority of basic skills training (n=11) (Figure 3). 9 courses used a combination of VR and dry lab simulation and 3 courses used just dry lab. In contrast, wet lab training alone was used for the majority of robotic procedural training (n=6). VR with wet lab and dry lab with wet lab were each used in 1 training course. 2 courses did not specify the training modalities used. A greater portion of training was delivered to more experienced participants in comparison to laparoscopy. Only 8 courses (31%) included medical students or interns (Figure 2.4, p. 35)

Open Surgery

22 studies offered open surgical training. 15 courses reported training for basic open surgical skills whilst 12 courses reported open procedural training. Apart from 1 course that used wet lab models, basic open surgical skills were taught entirely with dry lab models. In contrast, 6 procedural courses used only wet lab models, 2 used only dry lab models and 1 used both dry and wet lab simulators (Figure 2.3, p. 34).

All basic skills training was given to medical students, interns or junior residents. In contrast courses offering advanced open surgical training were provided mostly to junior residents (n=8) with the reminder open to medical students (n=2) and consultant surgeons (n=2) (Figure 2.4, p. 35).
2.3 Results

GI and GU Endoscopy

14 courses involved various training on endoscopy. 8 provide training on GI endoscopy, five on ureteroscopy and two on cystoscopy. 7 courses used a mixture of simulation techniques (combining VR, dry lab and wet lab or all three) whilst the remainder used either dry lab (3 studies) or virtual reality simulation (3 studies). 4 courses involved consultant surgeons or senior residents whilst 2 involved medical students. The remainder included only junior residents.

Microsurgery

Three courses on microsurgery skills training offered either basic and procedural training (n=2) or basic training alone. The basic skills courses used only dry lab training whilst the basic and procedural was provided through a combination of the dry lab and wet lab models or just wet lab. 2 courses recruited junior residents whilst 1 did not specify resident experience.

2.3.4 Teaching Techniques

131 studies reported sufficient detail for analysis of teaching technique. The majority of studies used more than one teaching technique (n=99, 76%). Knowledge-based training was provided in 59% (n=77) of courses. Regarding technical skills training, SDL and proctoring were used either alone or in combination in 73 (55%) and 95 (71%) courses respectively. Both observation and debrief were used much more infrequently. Observation was used in 15 courses, all of which involved procedural-based training for experienced participants (12/15 courses were for senior residents or consultant surgeons). Debriefing was used in 3 studies. In 2 of the studies debriefing was used for NTS training whilst in the remaining study it was used to provide technical feedback following suturing[140].
2.3.5 Training Time

Training time data were reported in 90 studies. Median training time was 16 hours (range 1-480 hours). There was no difference in the length of training between surgical technique (Kruskal-Wallis test p=0.5) (Figure 2.5, p. 38).

Fig. 2.5 Training Time by Surgical Technique (Median ± 95% CI)
2.3.6  Numbers of Participants

For 123 studies in which data were reported, total numbers of participants ranged from 2 to 477 with a median of 21. There were no significant differences in participant numbers between the surgical modalities (p=0.9).

2.3.7  Proficiency Based Training

75 courses (55%) used a proficiency-based training paradigm for some or all aspects of training. 3 studies did not report enough detail for appropriate analysis[147, 259, 213]. Various different standards for proficiency were used. 42 of the courses used proficiency standards based on expert performances of which 22 studies used the median or mean expert performance score. Other techniques for setting expert derived benchmark scores included 2 x standard deviations + average expert score (n=4)[231, 249, 145, 193] and expert score – 1SD (n=1)[250]. 9 courses used established scoring system; FLS (n=10), MISTELS (n=1) or the FSRS (n=1). 1 study did not report how the expert standard was set[175].

Time was used as the principle proficiency standard in 11 studies. Other approaches included error scores or error checklists with standard determined by experts (n=2)[219, 71], arbitrary pass marks (n=2)[228, 273] participant derived proficiency standards (n=5), the Angoff method (n=3) or subjective review by course instructors (n=2)[121, 176]. 7 studies used the inbuilt simulator metric benchmark scores to determine proficiency.

Training progression was determined by other, non-proficiency standards in 12 courses. These included self-determination by course participants (n=3) and a fixed period of training or repetitions (n=8).
2.3.8 Performance Assessment

The majority of courses (n=125, 91%) used an assessment tool for some or all aspects of training. 10 studies measured performance only by participant self-evaluation and 2 studies did not report a formal course outcome assessment. Global rating scales (GRS) were used in 53 studies. 21 studies used a procedure specific checklist and 48 studies assessed outcomes using computer measured metrics. Error based scoring systems such as the FLS were used in 38 courses. 10 studies relied on the unstructured assessment of experts (competent/non-competent or generic visual analogue scale (VAS) assessment). 20 courses used time as a major assessment indicator but only 6 courses used it as the sole criterion of outcome. Similarly, self-rating by participants was also used in 20 courses with 10 courses using this as the primary assessment tool.

Formal rater reliability assessment was undertaken in 27% (n=27) of studies in which it was indicated (studies using a GRS, procedure specific checklist or other operator dependent assessment tool). Inter-rater reliability ranged widely between 0.3 and 1. Mean (standard deviation) inter-rater reliability was $0.79 \pm 0.18$. There was a significant different between GRS and procedural scores (mean inter-rater reliability $0.72 \pm 0.19$ and $0.87 \pm 0.12$ respectively, $p < 0.005$).

2.3.9 Training Outcomes

Due to the significant heterogeneity in outcomes reported in the studies included in this review, pooled analysis was not undertaken. Outcomes were reported by 131 studies (96%). 83 studies showed positive results for all key outcome measures in those trials whilst 46 studies reported some negative findings but overall showed a benefit with training. 2 studies showed no benefit to training[179, 252].
2.3 Results

2.3.10 Quality Analysis

Quality assessment was possible in 132 studies. Mean (SD) MERSQI score (maximum score = 18) was 12.2 ± 2.8. In total 47 studies were considered high quality with a score greater than or equal to 14/18. Median MERSQI scores for each domain are shown in Figure 2.6, p. 42. There were a number of areas in which the studies scored badly. Most trials were single group with or without pretest/post test analysis (65%, n=86), most courses were held in single institution (75%, n=100) and 36% (n=48) of trials used unvalidated assessment tools. Outcomes were another area of weakness, 68% had a skills outcome whilst a further 13% had only opinions as an outcome. In contrast data type and data analysis were performed well (median score 3 for both fields).
Fig. 2.6 MERSQI Analysis showing median ± 95% CI Scores
2.4 Discussion

Unlike previous reviews of simulation training in surgery, this study takes the unique approach of focussing on the course structure and delivery of simulation for surgical education. The literature reporting outcomes and assessments of simulation training have been extensively and comprehensively examined. Numerous high-quality reviews have been published showing the effectiveness of simulation based training across various disciplines including robotic surgery[53, 215, 50, 59].

This review encompasses 137 studies reporting simulation training programmes for training surgical technical skills. Since the first studies were published in 1992, there has been an exponential increase in publication of studies reporting the outcomes of simulation training. Similar results have been reported previously[113]. Training has been reported from around the world although the USA, UK and Canada predominate. The majority of courses offered training in laparoscopy and despite the advances in VR, dry lab models remain the most frequently used by a considerable margin. Comparisons of different training modalities showed variations in how training was delivered. Overall dry lab simulation tools were most commonly used especially for basic training. In contrast a greater portion of procedural training was delivered with wet lab models. Even in laparoscopy, the predominant training modalities were dry lab and wet lab models. The only exception was robotic surgery where VR simulation was used for the majority of basic training. Course participation was divided relatively equally between groups. Junior residents comprised the largest group attending 42% (n=57) of courses. Medical students, senior residents and consultants attended 23% (n=31), 25% (n=34) and 19% (n=26) of course respectively.

Training structures and methodologies were investigated across the 137 studies. Most courses used a combination of technical and knowledge based or theoretical
training and lasted for 2 days (median training time 16 hours). Proficiency based training was used in the majority courses with the average expert score the most common proficiency standard. Most courses utilised a formal assessment method. GRS were used most frequently (45% of courses) while 41% used computer based metrics. Whilst other less objective measure of technical performance were used, only a minority of courses relied on them alone. Interestingly while procedural based checklists were used less frequently, inter-rater reliability was significantly higher than GRS (p<0.005). The overwhelming majority of studies reported positive training outcomes with only 2 (1%) failing to demonstrate a benefit for training. Study quality was assessed using the MERSQI framework for medical education literature. In total 34% (n=47) were high quality. A slightly higher proportion reported on laparoscopic skills training (n=31, 36% of all laparoscopic studies) whilst only 7 robotic studies were high quality (27% of all robotic studies). Studies scored highly for data collection and analysis but more poorly for study design, sampling and outcomes. Surprisingly 36% of studies used wholly unvalidated assessment tools.

This comprehensive review of the provision of simulation training as reported in the academic literature provides some useful insights. It is encouraging to see the increasing rate in publication of simulation studies although the proportion of high-quality studies remains low. This is comparable to previous reviews of simulation-based training[53, 86]. The courses themselves were overall relatively small scale, consisting of single centre programmes lasting on average 2 days and reporting on outcomes from a median of 21 participants irrespective of the surgical modality. Again this corresponds to previous reviews of the simulation literature in general[50]. In contrast to prior reviews, the proportion of medical students (23%) was much lower than prior reports; Cook et al. [50] reported that 37% of participants were students and Cook et al. [53] reported 45%. For procedural based training, experienced participants made up greater proportions
of participants. This is in keeping with the established literature. For more junior, inexperienced participants pure skills training is appropriate; more experience learners require whole task training[23].

The majority of studies did use performance standards and proficiency-based training paradigms. The advantages of competency-based training where practice is directed towards a set standard rather than being directed by time or number of repetitions is widely acknowledged in the literature[162, 173, 229]. Studies used predominantly valid methods for standard setting[86]. More arbitrary methods such as subjective expert decisions or predetermined standards were reported but these were in the minority. Importantly simulator-based metrics, set by the commercial developers of simulators often without specific details as to how they are formulated, were rarely used in isolation to determine proficiency.

This review highlights a number of key areas in which simulation training remains deficient. Overall training courses continue to be delivered as mass practice events, with training delivered at fixed time points and for fixed periods time. There is likely due to the logistical difficulties in delivery training. Greater provision of distributed simulation is required which can be tailored to the trainees clinical and educational requirements[23]. Outcome measures are important both to assessing training outcomes and providing accurate feedback for deliberate practice. 36% of studies used unvalidated assessment tools which needs to be improved. Furthermore, GRS were most commonly used yet only few studies detailed any attempts to provide rater training. GRS do provide greater flexibility and subjectivity to capture nuanced differences in performance however reliability is worse than procedural rating scales particularly with untrained raters[109]. Similar differences between procedural checklists and GRS were seen in reported rater reliability scores in this review. Course evaluation and feedback are essential elements for implementing simulation training. Outcomes were reported by 96% of studies this review. It is encouraging that training outcomes were positive in
most trials however the extremely low figure of 1.5% for negative studies does raise the possibility of publication bias. The outcome measures themselves were all of lower quality with 81% of studies reporting only skills or subjective outcomes and not any further measures of training impact. Previous reviews have also shown that outcomes on behaviours and patient effects remain poorly reported[53].

Regarding robotic surgery in particular, this review has highlighted a number of key areas of development. Studies reporting on robotic simulation training were of a lower quality and 46% (n=12) used proficiency-based training. Interestingly a greater portion of VR simulation was used for basic training and wet lab for procedural training. Study participation reflected the target audiences well with higher proportions of more experienced surgeons undergoing advanced training.

The principle limitations are firstly that the results are limited by the available. As discussed above, there was significant heterogeneity in the studies preventing pooled analysis. Furthermore, a significant proportion of studies had relatively poor methodologies limiting the conclusion that can be drawn. Finally, the unique approach of this systematic review to including only studies that adequately report the simulation course details means that there is the potential to miss studies. An exhaustive search strategy was used with 2 authors collecting data independently to reduce this risk.

2.4.1 Summary

This systematic review of the current state of simulation training for general surgery and urology has highlighted a number of important insights into training. Analysing the outcomes of 137 studies reporting on simulation training programmes a number of important factors are highlighted. Simulation courses continue to focus on short set periods of training lasting on average 2 days with training delivered to small numbers of participants. Dry-lab and wet-lab simulation modalities continue
to be used most commonly aside from robotic surgery in which VR simulation is most commonly used for basic training. There is widespread integration of proficiency-based training with the majority of studies using empirical methods for standard setting and objective assessment tool. Study quality remains low and further research is required to evaluate the role of distributed simulation training and the further effects of simulation training on behaviours and patient outcomes.
Chapter 3

Competency Based Training in Robotic Surgery: Benchmark Scores for Virtual Reality Robotic Simulation

3.1 Introduction

Since its introduction in 1991, the sophistication of VR simulators has grown, with ever improving fidelity and realism leading to ever greater utilisation. Numerous studies have supported the role of virtual reality training for robotic surgery[156]. As discussed in Chapter 1, VR simulators offer a realistic imitation of robotic surgery and provide objective performance assessment that can be used to monitor competence and progression. Limited evidence for the predictive validation of a number of VR simulators has been published[106, 47, 32, 55, 246]. One study has shown that VR training is directly linked to improved operative robotic performance[55]. Other studies have used surrogate markers to indicate competence, either dry lab or wet lab cadaveric simulation. Whether VR training is able to lead to competence in more advanced, procedure specific skills is yet to be established.
One of the key advantages of VR training is that it allows for the rationalisation and streamlining of learning. Replacing the highly stressful and expensive operating room is not only safer for patients and cheaper but allows a more conducive learning environment to be established[132]. VR simulation in particular allows training to be centred on the trainee around their other clinical and extracurricular commitments. Such an environment allows more personalised learning to take place whereby trainees can focus on developing the particular skills in which they are deficient rather than learning being dictated by case load and patient availability. Effective training requires more than just repetitive practice. A central concept for effective training is deliberate practice. Introduced by Ericsson deliberate practice is characterised by a highly structured, goal orientated approach to training. It is based on a number of key principles; motivated learners; repetitive performance of a particular task; well define objectives addressing relevant skills or topics; effective assessment with reliable data, informative feedback and performance evaluation[70]. Ericsson demonstrated that specialised training and feedback provide the optimum conditions for nurturing performance improvement. Furthermore it is hypothesised that deliberate practice is the key driving force in developing expert performance over both innate ability and extended experience[70]. Studies have demonstrated the effectiveness of deliberate practice and shown it be substantially superior to traditional methods of clinical training in a range of disciplines[150]. Deliberate practice is also often combined with mastery learning. This can be characterised as a competency-based training model in which skills and knowledge are rigorously tested in relation to the a set high standard beyond that of competency alone without any restriction on training time[148]. The aim is to perform uniformly high training outcomes although training time is expected to vary amongst participants. Mastery learning requires established, evidence based minimum standards, baseline assessment, targeted instruction, reassessment and progression based only attainment of the passing
standard. When performed correctly, mastery learning has been shown to be associated with higher outcomes than non-mastery learning[50].

A critical component of mastery learning and deliberate practice is accurate performance evaluation. Assessment before and after training is important to ensure that the necessary standards have been achieved and that training has been successful. Evaluation is also important for training in itself: feedback to learners helps to direct their learning, aids motivation and provides a standard against which progression can be checked. Feedback was identified in a review of clinical training as the most important feature for simulation based medical education[111]. Yet feedback needs to be understandable, relevant and usable for the trainees. A major advantage of VR simulators is the automated feedback. While such data can be useful to trainers and trainees, it is often only loosely related to clinical outcomes.

The Hands-On-Training (HOT) courses run by the European Association of Urology (EAU) provide structured VR training that aims to familiarise participants with the basic and more advanced skills required for robotic surgery. Currently there is no assessment of a participants’ progression during the course. Consequently, neither educators nor participants are able to objectively establish if the training has been effective. This study aims to construct a virtual reality assessment programme to be conducted during EAU HOT courses with clear benchmark scores of competencies to be obtained during the EAU HOT course.

3.2 Methods

Hands on training courses are regularly conducted during EAU conferences and symposia. A prospective, observational study was conducted recruiting candidates from nine HOT courses (EAU 2013, European Multidisciplinary Congress on Urological Cancers (EMUC) 2013, EAU 2014, EMUC 2014, EAU Section of
3.2 Methods

Oncological Urology (ESOU) 2014, EAU Robotic Urology Section (ERUS) 2014, ESOU 2015, ERUS 2015, Utrecht Robotic Symposium) between March 2013 and September 2015. Training programmes focussed on VR simulation training using the Mimic DV trainer. Course participation was open to all conference delegates regardless of experience or training. Standard demographic details and surgical experience were collected through a pre-course questionnaire that was distributed to all participants. Participant’s simulation exercise results were downloaded from the Mimic DV Simulators and compiled on an electronic database.

Across all nine courses, 16 different exercises were completed. For each individual exercise, the VR simulator provided performance data on a variety of performance metrics. These can be divided into general performance metrics, applicable to all exercises, (Time to Complete, Economy of Motion, Master Working Space, Instruments Out of View, Excessive Force, Instrument Collision) or task specific metrics (Blood Loss, Broken Vessels, Misapplied Energy Time, Missed Target, Dropped instruments). Task specific metrics were only applicable to certain exercises; for example, blood loss is only relevant to dissection section exercises. The Mimic VR simulator provides a summary score which is automatically generated on completion of an exercise based on the individual’s cumulative performance across all metrics. Analysis was limited to the general performance metrics and overall score so that all benchmark scores were applicable to all exercises.

To comprehensively assess a participant’s performance, a range of test exercises were selected encompassing basic and advanced skills. Suitable exercises were selected by a panel of expert robotic surgeons and surgical educators based on task complexity, the particular skill(s) assessed by the exercise and analysis of the previous HOT course participant results. To ensure valid analysis, exercises had to be performed at least 80 times to be included. Basic tasks were used to assess performance in generic, core robotic skills. It was hypothecated that the
majority of participants would be able to complete these exercises competently with relatively homogenous exercise scores. In contrast, advanced skill exercises were predicted to necessitate more complex, task specific, robotic skills. The greater challenge posed by these exercises was expected to result in greater variability and overall lower scores.

Benchmark scores were set using a criterion-referenced method based on expert scores. Given that the HOT course aims to provide initial training in basic robot skills, benchmark scores were sought to identify participants that had achieved a minimum level of competency rather than proficiency. Mean expert scores have been shown to be effective for establishing proficiency\cite{37}. Establishing expert standards remains difficult. On the basis of a thorough literature review of surgical learning curves, expert robotic surgeons were defined as having performed over 75 robotic cases independently\cite{1}. Nevertheless determining standards for clinical proficiency remains difficult since learning curves are both procedure and surgeon specific. As a result, an expanded dataset of expert performance data was used for this study from both expert performances during HOT courses and from the worldwide Mimic database. Results were combined to produce an overall mean expert score for each exercise. Experts from the Mimic database were recruited from 6 institutions in the US, France and Sweden.

Potential benchmark standards for competency were set at 60%, 75% and 90% of the mean expert score. These were modelled against participant outcome data to identify an appropriate standard (Figure 3.1, p. 53).

Through consensus opinion of the expert panel, a competency standard of 75% of mean expert score was found to set a suitable standard, based on the performance of novice (no robotic surgical experience), intermediate (1-74 robotic procedures performed) and expert participants (>75 robotic procedures performed). Retrospective analysis of the time to complete metric showed appropriate competency
Fig. 3.1 Comparison of Participant Performance with Benchmark Scores
rates when compared to potential alternative standards of 90% and 60% (Table 3.1, p. 55).
### Table 3.1 Comparative Competency Rates for Benchmark Standards of 90%, 75%, and 60% for Time to Complete

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Competency Rate for Benchmark Standard of 90% Mean Expert Score</th>
<th>Competency Rate for Benchmark Standard of 75% Mean Expert Score</th>
<th>Competency Rate for Benchmark Standard of 60% Mean Expert Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pick &amp; Place</td>
<td>21%</td>
<td>39%</td>
<td>57%</td>
</tr>
<tr>
<td>Camera Targeting 1</td>
<td>36%</td>
<td>38%</td>
<td>80%</td>
</tr>
<tr>
<td>Peg Board 1</td>
<td>36%</td>
<td>56%</td>
<td>100%</td>
</tr>
<tr>
<td>Thread the Rings 1</td>
<td>12%</td>
<td>24%</td>
<td>47%</td>
</tr>
<tr>
<td>Suture Sponge</td>
<td>13%</td>
<td>26%</td>
<td>33%</td>
</tr>
</tbody>
</table>
To test the suitability of the benchmark scores, comparisons were made between participants with differing levels of robotic surgical experience. Participants were divided into novice (no robotic surgical experience) and intermediate groups (1-74 robotic procedures performed). Their respective performances were compared against the benchmark criteria and expert performance scores.

3.2.1 Statistical Analysis

Initial data analysis demonstrated non-Gaussian distribution of scores with uniformly positively skewed data. In order to generate a normally distributed data set, logarithmic transformation was performed[25]. All further analysis was performed on this lognormal data. In each case the geometric mean was calculated to provide an accurate measure of central tendency. All calculations were performed using SPSS Version 22.0 (Armonk, NY: IBM Corp).

3.3 Results

3.3.1 Demographics

223 participants completed 1565 exercises over the course of nine HOT courses. Demographic details and surgical experience are shown in Table 3.2, p. 57. The HOT course cohort was composed of resident and attending urologists from 21 countries. Overall robotic experience was low. Residents had a mean $\pm$ standard deviation experience of assisting in $26.8 \pm 62.0$ cases and performing $1.4 \pm 3.8$ cases whilst attending urologists had experience of $19.3 \pm 44.5$ cases and $7.3 \pm 33.1$ cases respectively. Robotic simulation experience was equally low with 50.1% having no prior simulation experience. Previous analysis has shown that the main factors influencing the overall score of participants were age and prior robotic
Table 3.2 Demographic Details of HOT Course Participants

<table>
<thead>
<tr>
<th>Demographic Details</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Mean Years, SD)</td>
<td>38.4 ± 9.5</td>
</tr>
<tr>
<td>Gender (%)</td>
<td>76% Male: 24% Female</td>
</tr>
<tr>
<td>Level of Training</td>
<td>48% Resident: 54% Attending</td>
</tr>
<tr>
<td>Robotic Assistance Experience (mean no. of cases, SD)</td>
<td>21.3 ± 52.0</td>
</tr>
<tr>
<td>Laparoscopic Surgical Experience (mean no. of cases, SD)</td>
<td>68.5 ± 135.1</td>
</tr>
<tr>
<td>Proportion of Participants with Robotic Simulation Experience</td>
<td>53%</td>
</tr>
</tbody>
</table>

experience[75]. Each participant completed a mean of 7 exercises during a HOT course.

### 3.3.2 Identification of Assessment Tasks

Nine simulator exercises were performed >80 times (Pick & Place, Camera Targeting 1, Energy Dissection 1, Suture Sponge, Peg Board 1, Camera Targeting 2, Peg Board 2, Ring Walk 2). From these, five were selected as suitable assessment exercises; three basic level exercises (Pick and Place, Camera Targeting 1, Peg Board 1) and two advanced level exercises (Suture Sponge and Thread the Rings 1). The three basic exercises tested fundamental robot skills including endowrist manipulation, clutching, 3-D vision and camera control. As predicted results of participant performances showed relatively homogenous results with a major proportion of participants achieving high overall scores. Yet interestingly all three exercises had a pronounced dichotomy in scores, with a clear division between...
trainees (Figure 3.2, p. 59). Closer analysis revealed that while most participants started with these basic skill tasks, there was little repetition even following poor performance and hence no scope for development.
3.3 Results

Fig. 3.2 Participant Performance Analysis
Both advanced exercises assessed the more complex skill of suturing requiring needle driving in addition to competent execution of basic robotic surgical skills. Performance analysis showed far greater variability in scores and overall lower scores as expected with the more complicated tasks.

### 3.3.3 Development of Benchmark Criteria

The benchmark score for the minimum necessary standard to be achieved was set to 75% of the geometric mean expert performance. Scores were calculated for all general performance metrics (Time to Complete, Economy of Motion, Master Workspace, Instruments Out of View, Instrument Collisions, Excessive Force and Overall Score) for each of the five exercises. Mean participant scores were compared to the benchmark and expert scores in each case (Figure 3.1, p. 53). For basic tasks (Pick & Place, Camera Targeting 1, Peg Board 1), participant mean scores closely correlated with the benchmarks. Scores for the two advanced tasks (Thread the Rings 1, Suture Sponge) were more disparate. The key metrics for basic skills tasks were Time to Complete, Economy of Motion and Instruments Out of View. In contrast only Excessive Force and Instruments Out of View were successful measures of competency for advanced tasks.

Overall score provided a gross summary however its usefulness in assessing competence was limited especially in basic exercises where most participants met the competency standard. Mean participant score met or exceeded the benchmark in all but one exercise. Master Working Space was found not to be a suitable performance indicator with all participants exceeding the benchmark criteria.
3.3.4 Comparative Benchmark Assessment of Novice and Intermediate Participants

Benchmark scores offer appropriate targets for both novice and intermediate participants (Figure 3.3, p. 62). The majority of novice participants demonstrated competency in the basic tasks however fewer were able to meet the benchmark score for the advanced skills assessment tasks. As expected intermediate participants performed better in the majority of tasks across the six metrics but still fell short of the benchmark scores in the more difficult exercises.

Time to Complete

The increasing complexity of the exercises is clearly reflected in the time to complete metric with a progressive rise in the benchmark standard. Intermediate participants demonstrated competency in 3/5 exercises. A proportionally greater skills gap is seen with novice candidates with the more advanced exercises in particular.

Economy of Motion

Economy of motion exhibits a similar progression in difficulty across the five exercises. Benchmark scores become increasingly challenging for novice candidates unlike intermediate candidates who remain close to the benchmark standard and even surpass it in the suture sponge task.

Excessive Force

Excessive Force offers a greater challenge for inexperienced participants. Experts are able to maintain low scores across both basic and advanced exercises. In contrast, scores for both novices and intermediates deviate markedly with Thread
3.3 Results

Fig. 3.3 Comparison of Novice and intermediate Participant Scores to Benchmark Scores
the Rings 1 and Suture Sponge. Uniformly low scores in the basic skill tasks limits its application in these exercises.

**Instruments Out of View**

Notably this metric does not adhere to the pattern of rising task complexity seen in previous tasks; both novice and intermediate participants score poorly for Camera Targeting 1. For the two more complex tasks, intermediate and novice participants perform significantly worse.

**Instrument Collisions**

Basic task scores are uniformly low for novice and intermediate participants preventing effective differentiation of participants. With the increasing difficulty of the tasks, the participants’ scores rise exponentially compared to expert scores.

**Overall Score**

As a weighted composite score, the overall score was be expected to be equal across all tasks, and this is demonstrated by relatively uniform expert scores. All participants achieved competency in the basic skill exercises although there was a greater variation between intermediate and novice scores. In contrast novice participants failed to meet the benchmark for either of the advance exercises and intermediates only just met the standard for Suture Sponge.

### 3.4 Discussion

This study has demonstrated the feasibility of developing relevant benchmark scores for VR reality training, based on expert performance, to set clinically relevant targets for participants of varying levels of experience. Results from this
study will enable the development of an evidence based, clinically orientated VR simulation training programme for robotic surgery.

Historically surgical education has been based on repeated practice, with learning both contingent on and judged by case experience. Whilst concepts such as minimum case numbers remain ubiquitous, the value of competency-based training through simulation is increasingly being recognised[111]. Virtual reality simulators such as the Mimic DV trainers together with their VR training programmes have been widely validated and shown to be an important component of robotic training. However, the training structure in which such VR training takes place is integral to effectual training. Potential benefits of simulation training remain dependent on the trainee and their ability to learn. Individual factors, such as cognitive ability, motivation, perceived utility of training and self-confidence, may account for a significant proportion of the variation seen in training outcomes[38]. Hence ‘real-world’ results of simulation training will not necessarily match those from a highly focussed trial setting.

Objective assessment is needed to confirm to both trainers and trainees that the educational objectives of a training programme have been met. This study has demonstrated that benchmark scores, based on expert performance, set clinically relevant targets for participants irrespective of their experience during regular HOT courses. Surgical simulation ultimately aims to improve the clinical performance and outcomes at the time of surgery and post operatively. As discussed in chapter 1, evidence for the transfer validity for VR simulation in robotic surgery has been reported. It must be recognised that this evidence still rests on the assumption that effective training leading to competent performance in the simulation environment is undertaken. Establishing that competency in a task has been achieved entails both objective performance assessment and determining the standards against which trainees should be measured. VR simulators provide a wide and exhaustive range of data on an individual’s technical performance. Yet
without set benchmarks, such results remain abstract and unrelated to clinical outcomes. Determining competence is key to effective training but establishing credible, appropriate cut off scores remains challenging[268]. Furthermore, in the high stake’s evaluation of competency, assessment demands an evidence-based approach against impartial standards.

A variety of approaches to determining standards for training has been developed and evaluated in the medical education literature. The contrasting groups method is frequently used[86]. Participants are deemed to be competent or incompetent (or qualified or unqualified) by a group of judges based on prior evaluation. Performance scores of the particular test, often assessed using global rating tools or checklists are then reviewed and the score that best differentiates these two groups is selected as the standard[139]. The borderline group method employs a similar methodology with participants divided into 3 groups based on prior performance: pass, borderline and fail. The average performance of the borderline group is then used to set the assessment standard. These examinee-centred approaches to standard setting rely on the ability to accurately identify competent and incompetent participants. Expert-novice comparisons are often utilised but such crude approaches preclude a precise determination of the competency standard. As demonstrated in this study, determining marginal surgical ability remains relatively difficult especially in the absence of universal standards for assessing technical ability.

Alternatively a test centred approach to standard setting may be undertaken. The Angoff method being the second most commonly used methodology in surgical assessment[86]. In these methods expert judges review test items and determine the standard a examinee will need to achieve to demonstrate competency[20]. Such approaches are well suited to simulation-based assessment as they require complete standardisation of the task with minimisation of any non-surgical factors that may affect outcomes. However the application to a metric-based assessment
is complicated by the abstract and very detailed nature of the data. Simulators report with complete precision a wide variety of measure of physical performance such as the exact number of movements of instruments and the path length taken. This volume of data limits any effective judgments being made on the standard scores aside from the very basic measurements such as time to complete. It should be noted that such simpler metrics are often used for this reason. Furthermore test centred approaches are limited by the introduction significant potential for potential for subjectivity and bias\[117].

The alternative approach to criterion-referenced benchmark standards used in this study based on expert scores offers a number of advantages. Using the scores of expert surgeons as an indicator for competency associates simulation training to clinical benchmarks of proficient performance. Technical ability will even vary amongst expert surgeons however it can be safely assumed that all will be competent and some will be truly expert. As a result such a benchmark offers a reliable and conservative technical standard\[77]. In addition the scores represent a clinically relevant and acceptable standard for trainees to pursue. Finally this approach also offers better interpretation of the very comprehensive but relatively incomprehensible metric scores and enables a better global assessment of a participant’s technical performance.

Criterion-based assessment has been used in the past although infrequently with the mean expert score predominantly used as the standard with the average expert benchmark representing proficiency\[2] or “optimal performance”\[136]. In contrast, the HOT courses studied in this trial aim to provide basic training so that participants gain competency in basic robotic surgical skills rather than reaching proficiency. For this reason, the benchmark criterion was set as 75% of the mean expert score.

Metric based benchmarks offer a number of benefits to trainees. Specific goals help motivate participants and the immediate feedback highlight skill domains
that require improvement, aiding reflection and deliberate practice\cite{68}. Objective benchmark scores across different robotic skills exercises help promote modular training. Division of training into sequential tasks of increasing difficulty mirrors the process of motor skill acquisition\cite{73}. Trainees initially gain familiarity with robotic controls and basic skills, such as clutch control, camera control and endowrist manipulation. Subsequently, trainees apply these skills to more advanced tasks such as knot tying or suturing. This involves both refining their basic skills alongside learning such advanced techniques. Using benchmark criteria to govern progression ensures that course participants achieve competency in basic skills before progressing to more complex tasks. In contrast, unstructured training with progression regardless of scores risks poorer training outcomes. This may explain the poor progression of previous participants during HOT courses (Figure 3.2, p. 59).

A number of limitations to this study need to be considered. Use of expert scores to set the benchmark rests on a number of assumptions. Whilst there is substantial evidence to support the validity of VR simulators, the results are based upon appropriate training on the simulator to learn the necessary psychomotor skills. The risk of assessment through outcome metrics as used in this study is that participants will focus on improving their score rather than developing their technique. Although our use of a variety of metrics that evaluate different aspects of the technical performance will reduce this risk, there remains the potential for participants to learn only to complete the specific task rather than acquire the necessary robotic skills. Mentorship and teaching throughout the course are vital to avoid such training errors. Finally given the variations in metric data between different robotic VR simulation systems, benchmark scores are not transferable. As a result, benchmark scores need to be developed individually for each simulator. It has been shown that this method is generalisable with a similar methodology used to establish benchmarks for the RobotiX mentor simulator\cite{256}.
Within robotic surgery, progressive, stepwise simulation curricula have been proposed for robotic surgical training by various authors\cite{72}. Yet explicit performance assessment has not yet been implemented into any curriculum. This study has demonstrated the potential for using expert derived performances scores to develop reliable performance measures to aid the training and assessment of robotic surgical performance. On the basis of this a protocol for basic robotic surgical skills training using the Mimic VR simulator has been proposed. Further work needs to be undertaken to establish similar protocols for other robotic simulators and demonstrate the validity of this training model.

3.4.1 Summary

Analysis of the HOT course data has provided viable benchmark scores. Using an evidence based approach and extensive validation using a large database of VR simulation training data, this study has demonstrated that a benchmark of 75% below the mean expert score offers a challenging but obtainable score. This approach provides allows basic VR simulation training to be evaluated in the context of clinically relevant standards which will support participants and mentors completing basic robotic HOT courses. Continued analysis of HOT results will allow adaptation of these threshold values. Clear goals set through benchmarking offer objective targets for students and shift training from case volume-based training to a more efficient competency-based curriculum.
4.1 Introduction

VR training simulators originated with the Link trainer developed in 1929. This mechanical flight simulator allowed pilots, for the first time, to practice flying by instruments. Actual VR simulators were first introduced in the 1960s for flight training by the American military and NASA. It was not until 1993 that Satava [212] developed the first medical VR simulator; a rudimentary abdominal teaching tool that for the first time provided an interactive computer generated environment to aid medical training. Effective VR simulation training was feasible only because the necessary software was available to meet the two key requirements: accurate detail and interactive environments [212]. Subsequently a variety of experimental simulators were developed for laparoscopic, bronchoscopic and endoscopic surgery [238, 242, 243, 101]. These led to the development of the first commercially available simulators including the MIST-VR (Mentice, Gothenburg, Sweden) and The LapSim (Surgical Science, Gothenburg, Sweden) laparoscopic simulators. Since then, advances in computer hardware and software have con-
tinued to drive the development of ever more realistic and complex VR surgical simulators.

To date the majority of VR trainers especially for laparoscopic and robotic surgery provide basic skills training. These consist of core motor movements, techniques and skills. For example, in robotic surgery they principally consist of endoWrist manipulation, clutching, three-dimensional vision, dexterity, tissue handling, instrument control and camera control. Suturing is also often included in the latter stages of basic skills training. VR simulators provide often abstract tasks for example moving blocks from one post to another or guiding a ring over a wire without the ring touching the wire. One of the most successful basic skills training programmes is the Fundamentals of Laparoscopic Surgery (FLS) which has been most widely assessed[272]. Comprising e-learning, hands-on skills training, and an assessment tool it is used around for the world for training and assessment in basic laparoscopic skills. In robotic surgery a number of training programmes do include basic skills training to varying degrees[226, 6, 254]

More recently, further advances in software and hardware have enabled the development of increasingly realistic surgical VR simulations. Whilst VR simulation is acknowledge as being effective in teaching basic surgical skills, bridging the gap between such isolated skills training and undertaking full surgical procedures in the OR requires further extensive training. With the ability for high fidelity recreations of the surgical environment, VR simulations have been able to extend beyond abstract training tasks and recreate complete or parts of surgical procedures. Even for basic VR trainers, sophisticated software and hardware is required to accurately model complex elements such as shadows, collision effects, and topological changes due to tearing, grasping, cutting. With procedural VR, the complexity is greatly increased with the need to accommodate surgical factors such as the effects of instruments or sutures on tissues, physiological responses
such as bleeding and accurate anatomical modelling\cite{251}. The simulator must also be able to provide useful and objective assessments of performance.

As discussed in Chapter 1, six VR simulators offer procedural robotic simulation. The majority have undergone only rudimentary validation aside from the RobotiX mentor which offers the most extensive procedural training. A number of laparoscopic VR simulators offering procedural VR training have been evaluated. Studies have demonstrated that procedural VR training is more effective than no training\cite{131, 14}. In contrast compared to cadaveric animal training, laparoscopic procedural VR was found to be less effective\cite{247}. The only previous assessment of procedural VR training for robotic surgery was of the “Tubes” module for urethrovesical anastomosis training (UVA). This demonstrated significant training benefits when compared to no training as assessed on a synthetic, dry lab model\cite{123}. Even then, the module itself remains relatively basic in its design with a simple cylinder used to portray the urethra (Figure 4.1, p. 71).

![Image of Tubes 3 module](image)

Fig. 4.1 Example of Tubes 3 module (adapted from "Virtual Reality suturing task as an objective test for robotic experience assessment by Liss et al. [138]")
An alternative model for procedural training has been developed for the Mimic system using surgical video overlaid onto VR which provides both knowledge based and technical skills training. This composite system has been called augmented reality and has been shown to be effective for partial nephrectomy training\cite{107, 106}.

4.1.1 Objectives

This chapter describes the evaluation of the RobotiX VR simulator for full procedural training in robotic surgery. The objectives are to (I) assess the validity of the RobotiX VR simulator as a training tool and (II) compare the use of procedural and basic VR simulation for training robotic surgical skills.

4.2 Methods

A multi-institutional, prospective mixed-methods study was principally conducted in the Vattikuti Institute for Robotic Surgery, King’s College London and the Chitra Sethia Minimal Access Centre, University College Hospital, London. Data collection was conducted between the March and December 2016. Validation of the procedural training simulator was undertaken by a prospective, non-randomised comparative study. A randomised controlled trial was conducted to compare procedural against basic VR simulation training and no training.

At the time of writing, the RobotiX mentor (3D Systems, Airport City, 70151, Israel) offered 7 different procedural training modules covering gynaecological, thoracic, urological and general surgery. The radical prostatectomy module is the only module for urological surgery. Alongside video-based didactic training, the hands-on VR training can be undertaken either with or without step-by-step procedural guidance. Given the unfamiliarity of some participants with radical prostatectomy, guided training modules were used throughout the study. Three training tasks
form the RARP training module; bladder neck dissection, neurovascular bundle dissection and urethrovessical anastomosis. Since the neurovascular bundle training task required specific knowledge anatomical knowledge of successful completion, the authors decided to exclude it.

4.2.1 Pilot Study

A pilot study was undertaken whereby a robotic surgeon completed an extended period of training on the radical prostatectomy modules. The aim of the pilot study was to collect baseline data on the learning curve for procedural based training on the RobotiX mentor. Pilot study data would also be used to identify the appropriate metrics for the assessment the participants during the further trial.

4.2.2 Participants Selection and Initial VR Simulation Training

Novice participants were recruited by open invitation from London medical schools (King’s College London; Bart’s and The London School of Medicine and Dentistry; Imperial College School of Medicine; University College London Medical School). Surgeons were recruited from the Guy’s Simulation and Cadaveric Robotic Surgery Training Course’ Guy’s and St Thomas’ NHS Foundation Trust, the European Association of Urology (EAU) Section of Robotic Surgery Annual Conference, Milan and the EAU Annual Conference, Munich.

Participants were classified according to their robotic experience into novice, intermediate and expert groups. According to prior studies on learning curves for robot assisted radical prostatectomy (RARP), a significant improvement in outcomes is seen after 50 cases[1, 115]. Whilst not signifying completion of the learning curve, this was taken as the standard for differentiating expert and
intermediate surgeons in this study. Intermediate surgeons were defined as having performed between 1 and 49 RARP independently. Novices were defined as having no experience of performing RARP.

Prior to the validation study, all novice participants underwent basic robotic skills training (Figure 4.2, p. 75). All participants completed three FRS tasks during a one-hour training session (Ring Tower Transfer, Railroad Track, Vessel Energy Dissection). These tasks were selected to provide the exposure to core robotic skills including endowrist manipulation, camera navigation, dissection and diathermy use. Intermediate and expert participants were guided through the controls and permitted to perform the familiarisation tasks on FRS curriculum if they felt it necessary. No data was collected during this familiarisation training. Following familiarisation, all three groups completed the guided bladder neck dissection (BND) and guided urethrovesical anastomosis (UVA) tasks. Metric scores from the RobotiX Mentor were recorded.
Fig. 4.2 Flowchart for initial validation study
4.2 Methods

4.2.3 Validation of the Robot-Assisted Radical Prostatectomy Modules

Validity evidence for the effectiveness of procedural prostatectomy training using the RobotiX mentor was collected on content, internal structure and response process. Following completion of the guided BND and UVA tasks, intermediate and expert participants completed a quantitative and qualitative questionnaire. Demographic details and experience were recorded. The experts and intermediate level participant answered questions on the perceived realism of the simulator in comparison to the da Vinci robot using a 5-point Likert scale. The content and realism of the prostatectomy module was also assessed through questions regarding the module in general and each task specifically. Feasibility and acceptability were of procedural VR were also assessed.

The novice group continued on to perform a mean of 5.5hrs of further procedural training over 5 weeks consisting of 1-hour training sessions using the guided BND and guided UVA tasks. Performance metrics were recorded for each attempt. Participants were instructed to follow the guidance from the software programme on the successful completion of each task. Competency benchmarks were not available for the procedural training modules. Raw metric scores were used by the participants and the instructors to guide training.

Following training, the novice group completed a questionnaire on their demographic details and previous experience. Metric data from this training was used to analyse learning curve and training outcomes. Metric internal consistency was assessed using Cronbach’s $\alpha$. The study process is demonstrated in Figure 4.2, p. 75.
4.2 Methods

Internal Structure Assessment

The relationship between the metric scores and the proficiency of the novice participants was explored through internal structure assessment. Internal consistency of the 10 significant metrics was evaluated through analysis of the novice participants' scores on their second attempt. Cronbach’s $\alpha$ was used to measure consistency. Test-retest consistency was evaluated by comparing participant score for BND and UVA when completed immediately after each other. A two-way mixed model intraclass correlation coefficient (ICC) was calculated.

4.2.4 Randomised Controlled Trial of Procedural vs Basic Simulation Training

Prior to training all novice participants were randomised using a block randomisation protocol using an online randomisation tool (http://www.randomization.com). Participants were randomised into two groups using a block size of 4 to either procedural simulation or basic simulation. Figure 4.3, p. 78 provides an overview of the training protocol.
Fig. 4.3 Trial Design of Basic Vs Procedural Vs No. Training Randomised Controlled Trial.
4.2 Methods

The procedural training group completed the training programme as detailed in Figure 4.2, p. 75. The basic simulation training group underwent a parallel training programme. Firstly they completed the familiarisation tasks (Ring Tower Transfer, Railroad Track, and Vessel energy dissection). They then underwent further basic surgical skills training using the FRS curriculum and a continuous suturing module. Participants completed all six modules providing training on basic robotic technical skills principally utilising in-programme guidance. Successful completion of each module was determined by competency scores provided by the simulation programme. Additionally, ad-hoc training and guidance was provided by a study member and expert robotic surgeon to all study participants in both groups.

After 3 hours of training both groups underwent the first assessment of transfer of skills on fresh frozen cadavers. Performances were compared to a further group of novices who had no training. Both intervention groups then underwent a minimum of two further hours of training before a final dry-lab skills assessment.

Skills Transfer Assessment on Fresh Frozen Cadavers

Following the initial 5 week VR simulation training, all participants underwent identical skills assessment on fresh frozen cadavers within a simulated operating room. Cadavers were set up with an “Igloo” disseminated operating room to provide a realistic environment[31]. Given the training, RARP was selected as the assessment task and the cadaver was placed in the Trendelenburg position. This assessment aimed to evaluate the transfer of generic robotic skills developed through the skills training programme. In addition to the 26 intervention group participants, 9 control participants without any robotic simulation experience were recruited using the same methods as reported in Section 4.2.2.

All 35 subjects were allocated a 15 min assessment time slot according to their availability irrespective of their randomisation status. A Da Vinci Si surgical robot was used and each participant was given a short period of familiarisation.
For each assessment, the participants were guided through the steps of a robot assisted radical prostatectomy by two study investigators. Given the participants inexperience, general guidance on what tasks to perform were provided however proctoring or any specific technical instructions were not provided. Complex steps such as the urethrovesical anastomosis were avoided to enable fair analysis of all participants. All participant performances were recorded using the robotic laparoscopic camera. Video recordings for one participant in the basic training cohort were corrupted and their results were therefore excluded from further analysis. Technical performance was evaluated post-hoc using the GEARS assessment tool by one study investigator who had undergone extensive GEARS training and validation. The assessor was blinded to the participants’ identity or allocation status (see Appendix B for details of the rater training and validation).

**Skills Transfer Assessment on Dry Lab Training Models**

A second assessment of the transfer of training was conducted at the end of the simulation training programme. 20 participants (10 from each group) underwent the final assessment. 3 participants from both groups were unavailable. Assessment was conducted at Chitra Sethia Minimal Access Centre, University College Hospital, London using a Da Vinci S surgical robot. Participants first underwent a warm up task consisting of a thread-the-rings dry lab model (Loops and Wire, 3-Dmed (Franklin, Ohio, USA)). For the transfer of training assessment, UVA was tested using a synthetic model (3-Dmed (Franklin, Ohio, USA). The specific assessment task focussed on the participant’s ability to perform robotic suturing in contrast to the more general assessment of robotic surgical skill assessed on cadavers. Participants were first allowed a practice attempt before performing the task under assessment conditions. A 10min time limit for the assessment task was set. Guidance from two study investigators was provided during the warm up component only. Performances were recorded from the laparoscopic
4.2 Methods

robotic camera. Videos were scored post hoc by a study investigator blinded to the participants using GEARS and the robotic suturing checklist list[85, 91].

**Learning Curve Analysis**

Direct comparison of training effects was compared through learning curve analysis. Two tasks were selected from each training curriculum. For basic VR training, the Rail Road suturing and Ring Tower Transfer tasks were selected on the basis of the skill set requirements for successful completion. These were compared to BND and UVA. Common metrics between all four tasks were analysed. Across all tasks, analyses were undertaken to determine whether there was a significant change in scores over the training period. For those metrics in a significant change was identified, learning curves were plotted. Given participants completed a mean of 5.5 hrs of training, learning curve assessment was limited to the first 5 attempts. To allow comparison between the different exercises, Z-scores were calculated.

**4.2.5 Power Calculation**

The power calculation was based on a previous study evaluating procedural based VR training in robotic surgery using GEARS[48]. This study showed a difference in GEARS score of 2.5 points between procedural and standard VR training which we set as the minimum relevant difference for this study. On the basis of these results, for an $\alpha$ 0.05 (two-sided) and a power of 80% ($\beta=0.2$, $Z\alpha=1.96$, $Z\beta=0.84$, SD=2) we required at least 12 participants in each arm for the primary analysis of the RCT.

**4.2.6 Statistical analysis**

Performance metric data was retrieved from the simulator and analysed using SPSS (IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY: IBM Corp).
Normality assessment was performed on a sample of 8 metrics across BND and UVA. For each metric data were found to have non-Gaussian distribution and non-parametric analyses of the data were undertaken.

Pilot data was analysed by plotting and visually assessing learning curves. Outliers were identified and excluded using the Rout method, Q = 5%. Comparison of expert, intermediate and novice performances was undertaken using multivariate analysis of variance (MANOVA). Homogeneity of covariance matrices were assessed using Box’s M test. Homogeneity of error variances was assessed using Levene’s test. If Levene’s test was violated, a more conservative α level of p < 0.01 was used to determine significance for that for that variable in the univariate analysis[237].

Graphs were plotted to illustrate comparative outcomes. Learning curve analysis was undertaken using a linear mixed model for repeated measures. Additionally learning curves were plotted and visually assessed. The mixed models assessed change in metric over the first five attempts. Dependency of repeated measures was taken into account by including a random-intercept for each patient. Maximum likelihood was used as the estimation method. All graphs were produced with GraphPad Prism (version 7 for Macintosh, GraphPad Software, La Jolla California USA,).

4.3 Results

4.3.1 Pilot Study Results

The experienced robotic surgeon completed 32 repetitions of the UVA module over 5 weeks. Learning curves for all 27 performance metrics reported by the RobotiX mentor were plotted and analysed (Figure 4.4, p. 83 & Figure 4.5, p. 84). Significant learning effects were seen in 14 metrics (Figure 4.4, p. 83). Of these
10 were common to BND and UVA and were then used as indicative measures of learning for the further studies (Table 4.1, p. 85).

Fig. 4.4 Metrics Demonstrating Significant Training Effects
Fig. 4.5 Metrics Not Demonstrating Training Effects
Table 4.1 Common Key Performance Metrics Showing Significant Learning Effects

<table>
<thead>
<tr>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clutch Usage</td>
</tr>
<tr>
<td>Total Time</td>
</tr>
<tr>
<td>Instrument Collisions</td>
</tr>
<tr>
<td>Number of Times Instruments Out of View</td>
</tr>
<tr>
<td>Total Time Instruments Are Out of View</td>
</tr>
<tr>
<td>Number of Movements of Left Instrument</td>
</tr>
<tr>
<td>Number of Movements of Right Instrument</td>
</tr>
<tr>
<td>Path Length of Left Instrument</td>
</tr>
<tr>
<td>Path Length of Right Instrument</td>
</tr>
<tr>
<td>Distance Moved by Camera</td>
</tr>
</tbody>
</table>

4.3.2 Validation of the Robot-Assisted Radical Prostatectomy Modules

45 participants were recruited for the validation of the radical prostatectomy training modules; 13 novice, 24 intermediate and 8 expert surgeons. Table 4.2, p. 86 provides the basic demographic details.
Table 4.2 Participant Demographic Details (* experiences includes robot assisted nephrectomy, prostatectomy and cystectomy)

<table>
<thead>
<tr>
<th></th>
<th>Novice</th>
<th>Intermediate</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Of Participants</td>
<td>13</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>Mean Age (Range)</td>
<td>23 (21-34)</td>
<td>36 (30-50)</td>
<td>45 (35-62)</td>
</tr>
<tr>
<td>No. of Medical Students</td>
<td>13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No. of Residents</td>
<td>-</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>No. of Consultant</td>
<td>-</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Robotic experience* (Mean No. (Range))</td>
<td>4 (0-15)</td>
<td>1 (0-8)</td>
<td>-</td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assisted</td>
<td>0</td>
<td>33 (0-205)</td>
<td>2</td>
</tr>
<tr>
<td>Performed</td>
<td>-</td>
<td>6 (0-40)</td>
<td>610 (50-2720)</td>
</tr>
</tbody>
</table>
Survey Results

Qualitative survey results on the realism, feasibility, acceptability and perceived importance of the radical prostatectomy modules were collected from all intermediate and expert surgeons. Ratings for the graphics and appearances of environments modelled by the VR software were high with bladder, urethra and prostate close to 4/5 (Figure 4.6, p. 88). The realism of the simulator was scored similarly highly aside from the hand controls. Overall graphics were also scored at 3.5/5. A common comment from the surgeons despite the realistic appearance, was that the feel of the tissues remained unrealistic, “the tissues are too jelly like, needle floats which is less realistic”.

Likewise BND and UVA were scored 3.44/5 and 3.74/5 respectively (not shown). When asked to score the importance of individual tasks in training for RARP, both tasks were scored highly with bladder neck dissection 4.32/5 and urethrovesical anastomosis 4.6/5.

The intermediate and expert surgeons also reported a recognition of the importance of VR procedural training, both for the specific tasks and robotic training overall (Figure 4.7, p. 89).
Fig. 4.6 Participant Ratings of Realism and Anatomical Fidelity
Fig. 4.7 Participant Ratings of Realism and Anatomical Fidelity
4.3 Results

Learning Curves

Learning curves were plotted for BND (Figure 4.8, p. 91) and UVA (Figure 4.9, p. 92). Both exercises demonstrated training effects in a number of metrics. Improvements in movements and path lengths for left and right instruments were seen in addition to reductions in total time and to a lesser degree clutch usage. Instrument collisions and total time/number of times instruments out of view showed only limited training effects. Overall a more marked improvement across all 10 metrics were seen with the BND task (Figure 4.8, p. 91) compared to the UVA task (Figure 4.9, p. 92).

Comparative Performance Analysis of Novice, Intermediate and Expert Surgeons

The 10 common performance metrics (Table 4.1, p. 85), were used to evaluate the performances of all 45 participants on UVA and BND. One-way multivariate analysis of variance (MANOVA) was conducted to determine the effect of the three experience levels (novice, intermediate, expert) on the 10 common performance metrics. Box’s M was not significant for BND (p = 0.844) and UVA (p = 0.673).

On multivariate analysis, for BND there were significant differences between expert, intermediate, and novice participants in the 10 performance metrics, Wilks’ $\lambda = 0.37$, $p < 0.001$. Levene’s test was significant at $p < 0.01$ for five dependent variables (path length left, path length right, number of times instruments out of view, number of movements of right arm, total time instruments out of view).

Significant univariate effects were seen for nine metrics (including those confirmed by non-parametric analyses) (Table 4.3, p. 93)

MANOVA examining UVA also demonstrated a significant overall effect of the level of proficiency (novice, intermediate, expert) on performance metrics, Wilks’ $\lambda = 0.36$, $F (20, 120) = 3.96$, $p < 0.001$. Levene’s test was significant for six
Fig. 4.8 Novices Learning Curves for Bladder Neck Dissection
Fig. 4.9 Novices Learning Curves for Urethrovesical Anastomosis
### 4.3 Results

Table 4.3 MANOVA for Bladder Neck Dissection Task

<table>
<thead>
<tr>
<th>Metric</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.006</td>
</tr>
<tr>
<td>Clutch Usage</td>
<td>0.385</td>
</tr>
<tr>
<td>Path Length Left</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Path Length Right</td>
<td>0.008</td>
</tr>
<tr>
<td>Instrument Collisions</td>
<td>0.04</td>
</tr>
<tr>
<td>Number of Times Instruments Out Of View</td>
<td>0.001</td>
</tr>
<tr>
<td>Number Of Movements Of Left Instrument</td>
<td>0.001</td>
</tr>
<tr>
<td>Number Of Movements Of Right Instrument</td>
<td>0.001</td>
</tr>
<tr>
<td>Distance By Camera</td>
<td>0.02</td>
</tr>
<tr>
<td>Total Time Instrument Out Of View</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 4.4 MANOVA for Urethrovesical Anastomosis Task

<table>
<thead>
<tr>
<th>Metric</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.04 (NS)</td>
</tr>
<tr>
<td>Clutch Usage</td>
<td>0.25</td>
</tr>
<tr>
<td>Path Length Left</td>
<td>0.008</td>
</tr>
<tr>
<td>Path Length Right</td>
<td>0.005</td>
</tr>
<tr>
<td>Instrument Collisions</td>
<td>0.06</td>
</tr>
<tr>
<td>Number of Times Instruments Out Of View</td>
<td>0.20</td>
</tr>
<tr>
<td>Number Of Movements Of Left Instrument</td>
<td>0.003</td>
</tr>
<tr>
<td>Number Of Movements Of Right Instrument</td>
<td>0.01</td>
</tr>
<tr>
<td>Distance By Camera</td>
<td>0.16</td>
</tr>
<tr>
<td>Total Time Instrument Out Of View</td>
<td>0.07</td>
</tr>
</tbody>
</table>

metrics (Time, Clutch usage, path length right, instrument collisions, number of times instruments out of view, number of movements of right instrument).

Univariate analysis, using a more conservative $\alpha$ for those variables that violated Levene’s test, demonstrated 4 significant metrics as highlighted in Table 4.4, p. 93.

Alongside these statistically significant variables, comparative performances were also assessed. Visual analysis of the UVA task demonstrated evidence of appropriate variations in performance between novice, intermediate and expert participants in the majority of metrics. Kruskal-Willis analysis was used to test for
4.3 Results

differences between groups. Experts completed the task in a shorter time (p=0.02) with less instrument collisions (p=0.1) and greater clutch usage (p=0.03). There was a tendency toward fewer movements in the left (p=0.4) and right (p=0.5) arms but differences were not significant. Actual path lengths did not differ substantially. Experts moved the camera further (p=0.004) whilst novice surgeons kept instruments in view both in terms of time (p=0.002) and number of events (p=0.001)(Figure 4.10, p. 95).

For BND, differences between experience levels was seen but to a lesser extent. Experts completed the task in less time (p=0.02), with greater clutch usage (p=0.001), fewer instrument collisions (p=0.04) but greater number of times instruments out of view (p=0.002) and moved the camera more (p=0.01). In contrast to UVA, number of movements and path lengths for left and right arms increased with experience (p>0.05) (Figure 4.11, p. 96).

**Learning Curve Analysis**

A linear mixed model was used to evaluate the performance of novices following either basic or procedure training. Change in metric scores over the first five attempts by each participant was analysed (Table 4.5, p. 97). Total time, number of movements of left instrument, number of movements of right instrument and path length of left instrument showed significant improvements across all four exercises. Two further metrics, clutch usage and instrument collisions, demonstrated significant improvements in the majority of metrics.
Fig. 4.10 Comparative performances of novice, intermediate and expert participants for UVA (median ± 95% CI)
4.3 Results

Fig. 4.11 Comparative performances of novice, intermediate and expert participants for BND (median ± 95% CI) for BND.
### Table 4.5 Linear Mixed Model of Novice Training on Rail Road, Ring Tower Transfer, BND and UVA

<table>
<thead>
<tr>
<th>Metric</th>
<th>Rail Road</th>
<th>Ring Transfer</th>
<th>Tower</th>
<th>Bladder Dissection</th>
<th>Neck</th>
<th>Urethrovvesical Anastomosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>p value</td>
<td>&lt;0.001***</td>
<td>&lt;0.001***</td>
<td>&lt;0.001***</td>
<td>&lt;0.001***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Time</td>
<td>0.02*</td>
<td>0.004**</td>
<td>0.04*</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clutch Usage</td>
<td>0.16</td>
<td>&lt;0.001***</td>
<td>0.10</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance Moved by Camera</td>
<td>0.02*</td>
<td>0.008**</td>
<td>0.003**</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrument Collisions</td>
<td>&lt;0.001***</td>
<td>&lt;0.001***</td>
<td>&lt;0.001**</td>
<td>&lt;0.001***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Movements of Left Instrument</td>
<td>0.02*</td>
<td>&lt;0.001***</td>
<td>0.10</td>
<td>0.04*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Movements of Right Instrument</td>
<td>0.08</td>
<td>&lt;0.001***</td>
<td>0.008**</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Times Instruments Out of View</td>
<td>0.14</td>
<td>0.01*</td>
<td>0.02*</td>
<td>0.02*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Time Instruments Out of View</td>
<td>0.04*</td>
<td>&lt;0.001***</td>
<td>0.02*</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Path Length of Left Instrument</td>
<td>0.03*</td>
<td>&lt;0.001***</td>
<td>0.03*</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Path Length of Right Instrument</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Further analysis of the learning curves of the metrics for these tasks shows differences between the FRS and procedural training modules (Figure 4.12, p. 99). Overall there is a noticeable improvement in scores between the two training techniques particularly in the first 2-3 attempts. Total time instruments out of view demonstrates little learning effects after the first attempt in any exercise. For Ring Tower Transfer likewise little improvement after the first attempt is seen in the number of movements in left or right instrument, and path length of left and right instrument. Neither Ring Tower Transfer or Ring Railroad track show substantive improvement after the first attempt for distance moved by camera or to a lesser degree total time. Greater training effects are seen in BND and UVA tasks especially for total time, number of movements for right instrument and path length for right instrument.
4.3 Results

Fig. 4.12 Learning Curves for Basic and Procedural Training
4.3 Results

Table 4.6 Test - Retest Correlation Analysis

<table>
<thead>
<tr>
<th>Metric</th>
<th>Test-Retest Correlation BND (ICC)</th>
<th>Test-Retest Correlation UVA (ICC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clutch Usage</td>
<td>0.37</td>
<td>0.86</td>
</tr>
<tr>
<td>Total Time</td>
<td>0.57</td>
<td>0.77</td>
</tr>
<tr>
<td>Instrument Collisions</td>
<td>0.83</td>
<td>0.71</td>
</tr>
<tr>
<td>Number of Times Instruments Out of View</td>
<td>0.47</td>
<td>0.64</td>
</tr>
<tr>
<td>Total Time Instruments Are Out of View</td>
<td>0.17</td>
<td>0.95</td>
</tr>
<tr>
<td>Number of Movements of Left Instrument</td>
<td>0.85</td>
<td>0.65</td>
</tr>
<tr>
<td>Number of Movements of Right Instrument</td>
<td>0.29</td>
<td>0.68</td>
</tr>
<tr>
<td>Path Length of Left Instrument</td>
<td>0.88</td>
<td>0.75</td>
</tr>
<tr>
<td>Path Length of Right Instrument</td>
<td>0.27</td>
<td>0.83</td>
</tr>
<tr>
<td>Distance Moved by Camera</td>
<td>0.69</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Internal Structure Assessment

The 10 significant metrics for BND and UVA (Table 4.1, p. 85) demonstrated a good level of internal consistency, as determined by a Cronbach’s \( \alpha \) of 0.77 and 0.68 respectively[125]. Test-retest analysis was also undertaken using the intraclass correlation coefficient (ICC) (Table 4.6, p. 100).

4.3.3 Skills Transfer Assessment on Fresh Frozen Cadavers

25 participants, having completed a mean of 5.5 hours VR training on the RobotiX mentor underwent skills transfer assessment on fresh frozen cadavers. Performances were compared to 9 control participants who had not undergone any previous robotic surgical training of any form. Blinded GEARS assessment of all performance videos was undertaken.

VR training (procedural or basic training) resulted in a significantly higher GEARS score than no training, mean GEARS score 11.3 ± 0.58 vs 8.8 ± 2.9 \( p = 0.002 \) (Figure 4.13, p. 101). When analysed separately, there were significant
differences between no training, basic VR training and procedural VR training; mean GEARS 8.8 ± 1.4 vs 10.7 ± 2.8 vs 11.9 ± 2.9 respectively, p = 0.03 (Figure 4.14, p. 102).

![Box plot comparing GEARS scores between control and VR training](image)

**Fig. 4.13 Comparison of any VR training vs no Training**

### 4.3.4 Skills Transfer Assessment on Dry Lab Training Models

Following a further two hours of either basic or advanced VR training, 20 participants (Basic VR training = 10; advanced VR training = 10) underwent dry lab assessment of a UVA. Procedural VR training resulted in a significantly higher suturing checklist score; mean score 14.7 ± 3.5 vs 10.5 ± 4.2, p = 0.03 (Figure 4.15, p. 103). GEARS scores were also higher following advanced VR training but
Comparison of Advanced, Basic and No Training

Fig. 4.14 Comparison of Basic, Procedural and No VR Training
the differences were not significant (16.7 ± 3.1 vs 14.0 ± 3.8, p = 0.1) (Figure 4.16, p. 104).

Fig. 4.15 Dry Lab Skills Transfer Assessment: Robotic Suturing Checklist Score

4.4 Discussion

This study provides further objective evidence to support the use of VR simulation training for robotic surgery. In addition, this study reports the first validation of procedural VR simulation training for robotic surgery. Validity evidence was
Dry Lab Skills Transfer Assessment: GEARS Score

Fig. 4.16 Dry Lab Skills Transfer Assessment: GEARS Score
collected for support a number of aspects of the construct validity for BND and UVA procedural training modules on the Robotix Mentor VR Simulator have been reported.

Principally this study has shown that a programme of VR simulation training (either basic or procedural) resulted in significantly better operative surgical performance as assessed in a simulated wet lab environment. Furthermore, procedural training resulted in better performances outcomes in comparison to basic skills training. Assessment using a specific technical task (UVA) also demonstrate higher scores for procedural over basic skills training however results were not significant.

Successful implementation of any new simulation training tool requires appropriate construct validation. To a certain degree this has been demonstrated for the BND and UVA procedural training modules on the Robotix Mentor VR Simulator. Limited evidence is available for its content validity. According the developers, 3d Systems, the training modules were designed in conjunction with expert robotic surgeons (3D Systems RobotiX Mentor, 2019). However, it is beyond the scope of this study to assess the validity of this evidence. Qualitative feedback from expert surgeons suggested that the tasks were important for robotic training and provided a realistic training environment (Figure 4.6, p. 88, Figure 4.7, p. 89). A number of assessments of the response processes for the training modules were undertaken. Both the individual tasks and the training modules as a whole were recognised by participants of all levels as important training tools. This corresponded with learning curve analysis from the assessment tasks which demonstrated appropriate patterns of skill acquisition over the training course (Figure 4.4, p. 83).

Internal structure of the training metrics provided by the RobotiX Mentor was demonstrated in a number of areas. Since metrics are produced automatically by the software’s algorithms, rater assessment could not be performed. Moderate internal consistency of the 10 relevant metrics (Table 4) was found and there was good test retest reliability for the UVA anastomosis task (Cronbach $\alpha$ 0.64-0.95).
In contrast test retest reliability of the BND task was much poorer (Cronbach $\alpha$ 0.27-0.88).

A number of assessments of the relationship between simulator performance and other variables was undertaken. The procedural training modules could discriminate between novice, intermediate and expert participants in a number of metrics including time to complete, clutch usage, time out of view and instrument collisions.

The principle findings from this study relate to consequences of training. The effects of training both on performance on the VR simulator as assessed through learning curves and the transfer of learning effects were tested. Learning curve analysis showed that training over time resulted in improvements in performances on the VR procedural training modules (Figure 4.8, p. 91, Figure 4.9, p. 92). These findings were supported by the results of the transfer of learning assessment. These demonstrated significantly better technical skills outcomes during cadaveric wet lab assessment in comparison to novice surgeons with either basic training or no training experience.

Using a randomised design, this study has reported the first data supporting the transfer of skills following procedural VR simulation training. Practice using the procedural training modules produced significantly better results in the skills transfer task compared to both no training or basic training. This study firstly provides further justification for the use of a structured simulation training programme for robotic surgery. Both training groups demonstrated better generally robotic surgical proficiency to the control group. The cadaveric task was developed to provide an accurate assessment of operative robotic skill and, unlike many previous studies, neither the intervention or control participants had any previous experience of the task. Procedural based training was further evaluated using a procedure specific task, UVA on synthetic model, which also found higher skills scores for procedural training. The results of this study have important
implications for surgical training. Even in novice participants, procedural training appears to enable better skills acquisition than basic skills training, which has not, the authors knowledge, been previously tested. The reasons for this are not clear. Anecdotal feedback from participants was that the procedural training was more enjoyable than the abstract basic skills tasks. Such more clinically relevant training may help to motivate participants and focus their training.

A number of limitations to this study need to be considered. Regarding the validation of the procedural training models, there were a number of discrepancies in the metric scores. Learning curves show that training by novices resulted in fewer movements of the camera and clutch. In contrast expert surgeons made more movements of the camera and used the clutch more. Experts also used instruments out of view while still having few instrument collisions. While the novice surgeons appeared to become more efficient, their techniques may have been flawed. They learned to complete the task “better” rather than improving their robotic technique. This highlights the importance of ongoing proctoring even when using VR simulation training programmes to prevent learning bad habits. Perrenot et al. [178] et al reported similar findings. Time is a popular index of surgical technical ability, but it does not provide any information as to how the task was executed[142, 156]. A novice may achieve a low time but this may simply be as a result of a rushed and inaccurate performance.

Secondly the study focussed on novice participants with smaller groups of intermediate and expert surgeons. Further investigation of the training effects of procedural training for more experienced trainees is required. Frequent feedback from the experts was that the procedural models were too structured. The surgeons attempted to perform the procedure according to their own technique however this could not be accommodated by the simulator and as a result they were penalised. Although the simulated environments maintain a high degree of fidelity, there are obvious limits to what can be performed and they remain
inferior to real life be that cadaveric, live porcine or OR experience. Procedural training therefore remains limited to more junior trainings who require training in basic technical and procedural skills.

Finally the tasks were guided by the simulation programme which expected a specific technique. Experts and some intermediates who were familiar with bladder neck dissections did not believe that the guided tasks were very accurate and were restrictive compared to the way that they currently practice that aspect of a RARP. The validation of the procedural training modules was also limited by a relatively small number of intermediate and expert participant who did not complete the extended training programme. Further research needs to be undertaken with more experienced surgeons, to determine the role of procedural VR simulation training further along the learning curve. Finally, although transfer of training was demonstrated, further studies need to be undertaken to demonstrate whether this benefit is also carried into the OR.

### 4.4.1 Summary

This mixed study method aimed to both validate a newly developed radical prostatectomy VR simulation training module and evaluate the transfer effects of procedural VR training. To date, procedural training remains very limited and is not evident from the literature what the exact role for procedural VR is in robotic simulation. Using a cohort of novice intermediate and expert surgeons, evidence to support a number of areas of the construct validity of the procedural VR simulator has been reported. In addition, the study has also demonstrated the transfer effects of procedural VR training to both wet-lab and dry lab tasks. The results of this study provide firm initial results to support the use of procedural VR for robotic surgical training. The study has shown evidence to support this novel procedural VR module as a valid training tool and a beneficial addition to a robotic
surgery training curriculum. Before this module can be used to assess individuals, however, validation from further centres is required, and an expert consensus on which metrics are important and accurately reflect a good performance need to be established. Continued studies looking at predictive validity would increase the evidence that procedural VR is an effective method of training surgeons. Continued development of this already state-of-the-art module could improve functionality and allow a smoother transition into surgical curriculum.
Chapter 5

A Comparative Analysis of Virtual Reality and Dry-Lab Robotic Simulation Training

5.1 Introduction

During the initial phase of the learning curve, both virtual reality (VR) and dry lab simulators are routinely used to provide technical skills training. Particularly within laparoscopic and robotic surgery, VR simulation is widely used as a principle component of the surgical curriculum particularly for early stage trainees. As has been discussed in chapters 1 and, when used appropriately VR simulation offers numerous benefits to the trainee surgeon and his trainer. Similarly, dry-lab simulation training is widely used to support basic skills training although it is more widely applied to laparoscopic than robotic surgery (see Chapter 2, p. 25). There is evidence to support the use of dry-lab and VR simulation for more advanced procedural training, although wet-lab models are more commonly used[1, 156]. The optimal role for each constituent component within a simulation curriculum remains unclear and managing the transition from basic to intermediate and then advanced level training remains poorly elucidated.

Both VR and dry lab simulation have been shown to be equally effective for laparoscopic training but high capital costs for VR simulators limits their
application in routine training. In comparison, the cost and availability of dry lab simulation equipment allow it to be more easily applied to training [112, 164]. Health economic analyses are more difficult for more advanced training which require proctoring. In contrast to laparoscopic training, cost calculations for robotic surgery vary considerably. VR simulators remain expensive, however in comparison to the cost of a Da Vinci system they offer a relatively cheap and readily available training alternative[205].

With the developments in VR technology for robotic training, more advanced training modules such as suturing are now available across all commercially available simulators (see Chapter 1, p. 1). In spite of this, no comparison has yet been undertaken between VR robotic training and the “real-world” environment of dry-lab training. There is evidence from laparoscopic simulators to suggest that whilst VR simulator offers a high degree of realism and is beneficial to skills training, differences remain between VR and dry lab simulators. This study aims to compare VR simulation with dry lab simulation in robotic suturing training.

5.2 Methods

This prospective, comparative study was conducted at two institutions; Centro Oncologico Fiorentino, Florence, Italy and King’s College London, UK. 43 novice robotic surgeons were recruited. No participants had any robotic surgical or robotic simulation experience. All participants were provided with information on the study and completed a consent form. Demographic details and operative experience were collected using questionnaires. Participants were allocated to either VR simulation training or dry-lab simulation training. Initial allocation was according to centre with training in Florence limited to dry lab simulation. Participants in London allocated to either dry lab or VR using simple randomisation (Prism
5.2 Methods

Version 6 (GraphPad Software, La Jolla, CA, USA). Available All 43 participants completed the study.

5.2.1 Simulation Training Tools

Dry lab simulation training was provided using the Da Vinci S and Da Vinci Xi surgical systems (Intuitive Surgical, Sunnyvale, USA). Training was performed within an insufflated abdomen model (Intuitive Surgical, Sunnyvale, USA). Initial training on fundamental robotic skills was delivered using two basic skill exercise models; the Sea Spike Pod (The Chamberlain Group, Great Barrington, MA, USA) and Pea on a Peg (3-DMed, Franklin, OH, USA). Following this the participants were taught robotic suturing; closure of a vertical incision using a foam suture pad (Limbs & Things Ltd., Bristol, UK). The final assessment ask was completion of a urethrovesical anastomosis on a synthetic training model (3-DMed, Franklin, OH, USA).

Virtual simulation training was conducted using the da Vinci Surgical Simulator (dVSS) “backpack” in conjunction with the Da Vinci Xi robotic console (Intuitive Surgical, Sunnyvale, USA). The dVSS is widely validated and runs the Mimic MSim VR software (Mimic Technologies Inc., Seattle, WA, USA) which offers a variety of training modules primarily in basic robotic skills[19]. Two exercises, Pick and Place and Matchboard 2, were selected to provide equivalent training to that given to the dry-lab simulation group. Mirroring the dry lab simulation training, this task was used by the surgical mentors to teach the participants basic robotic skills.

Both cohorts were assessed using Global Evaluative Assessment of Robotic Skills (GEARS)[85]. This six point scoring checklist has been shown to be equivalent to the performance metrics produced by the VR simulator and allowed direct comparisons to be made between the two groups[201]. During each training
session, all participants were assessed during the last 10min of each session by two robotic training experts and the mean of the two scores was recorded.

5.2.2 Simulation Training Protocol

All participants underwent equivalent training using either the dry-lab or virtual reality simulators. Each participant completed a total of three hours of training during which time they were mentored by an expert robotic surgeon (Figure 5.1, p. 114).

In accordance with the principles of simulation training[36], during the first training session participants were taught basic robotic skills; camera targeting, endowrist manipulation and clutching. In the second training session subjects were taught robotic suturing including needle driving and knot tying. Both groups were taught to suture a vertical defect. During the third training session participants underwent further suture training in their specific modalities before completing the final assessment exercise. All participants performed a urethrovesical anastomosis on a synthetic training model (3-DMed, Franklin, OH, USA) using the Da Vinci surgical robot. A time limit of 20min was set. Following each training session, the participant’s technical performances were rated using GEARS. All trainees completed a post-training qualitative survey.

5.2.3 Power Calculation

The power calculation was based on the results of the Cochrane review[164]. This study reported a standard mean difference between virtual reality and dry lab simulation training of 1.46. On the basis of these results, for an $\alpha$ 0.05 (two-sided) and a power of 80% ($\beta$=0.2, Z$\alpha$=1.96, Z$\beta$=0.84, SD=2.3) we required at least 20 participants in each arm for the primary analysis of this comparative study.
5.2 Methods

Fig. 5.1 Study Flow Diagram
5.2.4 Statistical Analysis

Data were normally distributed. Between cohort differences were compared using an independent sample T-Test. Repeated measures analysis of variance (ANOVA) was performed to test for significance in mean differences between VR and Dry–lab simulation across the training programme. Data were analysed graphically and means were compared at each time point using independent sample T Tests. To assess the robustness of the data, sensitivity analyses was performed comparing outcomes from all participants to only those participants randomised at the London centre. Data was analysed using IBM SPSS version 23 (Armonk, NY, USA). Graphs were created using Prism Version 6 (GraphPad Software, La Jolla, CA, USA).

5.3 Results

The mean age of the participants was 23.5 years and 60.5% were male. Medical students formed the majority of participants (93%, n=40) Surgical and simulation experience between the groups was equally low (Table 5.1, p. 116). Neither group had experience in performing surgery independently. All participants were assessed by two raters with the mean GEARS score taken. Raters demonstrated a high degree of agreement with an interclass correlation coefficient of 0.78[129].

As expected both training modalities demonstrated beneficial training effects with rising GEARS scores over the three sessions (Figure 5.2, p. 117). As a result of the training, mean GEARS scores increased by 5.6 and 2.5 following dry-lab and VR simulation training respectively. This difference in training outcomes was confirmed by repeated measures ANOVA. An overall significant difference in training effects was found between the two training modalities across the three
training sessions (p = 0.034). In addition, a progressive rise in GEARS score is seen with dry-lab training; VR training resulted in a lesser rise in GEARS score.
5.3 Results

Fig. 5.2 Comparative Performances Following Dry Lab and VR Simulation Training
5.3 Results

Table 5.2 Technical Performance Scores following Dry Lab and VR Training

<table>
<thead>
<tr>
<th></th>
<th>Dry Lab Simulation</th>
<th>VR Simulation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean GEARS Score ± SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session 1</td>
<td>10.5 ± 0.5</td>
<td>11.8 ± 0.5</td>
<td>0.06</td>
</tr>
<tr>
<td>Session 2</td>
<td>12.0 ± 0.7</td>
<td>12.8 ± 0.5</td>
<td>0.37</td>
</tr>
<tr>
<td>Session 3</td>
<td>16.1 ± 0.5</td>
<td>14.3 ± 0.6</td>
<td>0.03</td>
</tr>
</tbody>
</table>

On individual analysis of session outcomes, the differences in scores in training only reached significance after the third training session when the more complex UVA task was assessed (Table 5.2, p. 118).

Overall the training was well received and the educational impact was clearly recognised. The post course survey reported a mean course satisfaction score was 5/5 with all participants endorsing the course. When the training modalities were compared, dry-lab trainer achieved marginally higher satisfaction scores than VR simulation (4.2/5 vs 4.9/5 respectively).

5.3.1 Sensitivity Analysis

Limiting analysis to only those participants who were randomised in the London centre did not demonstrate significant variations in outcomes. Direct comparisons of performances during the three tasks revealed equivalent results to the entire cohort although results for the first and third training session did not meet significance (Figure 5.3, p. 119). Repeated measures ANOVA demonstrated both a significant increase in GEARS score over the three training sessions and a significance difference in scores across the sessions between the two training modalities. A greater increase in GEARS score was seen with dry lab training with a mean final GEARS score 15.2 compared to 14.3 for VR training confirming the primary results of this trial.
Fig. 5.3 Comparative Performances Following Dry Lab and VR Simulation Training: Randomised Participants Only
5.4 Discussion

This comparative study of VR and dry lab simulation training for robotic surgery has shown that, whilst both VR and dry lab simulators were effective training tools, dry lab simulation training resulted in a greater improvement in technical skill. However, this difference only became apparent following the third training session on the urethrovesical anastomosis suturing task. Following the first two training sessions, the VR simulation achieved higher GEARS scores than the dry-lab group. Despite the identical console used for the VR and dry lab training, dry-lab trainees appeared to initially find the basic robotic tasks more difficult. Yet in a short space of time, greater improvement was seen with dry lab training.

As discussed above, simulation training has undergone extensive validation both in laparoscopic and robotic surgery. Both have been shown to effective in developing surgical skills[1]. Yet a debate continues over the form and structure of simulation training. Wet lab training is primarily focused toward more advanced, procedural skill whilst dry-lab and VR simulation offer basic surgical skill training. Meta-analysis of VR and dry lab simulation training outcomes for laparoscopic surgery have demonstrated that whilst both can be effective for training, they may not necessarily be interchangeable[271]. Basic tasks can be taught equally well using either modality. With more experienced participants or more complex tasks, conclusions are less clear[248]. A number of studies in laparoscopic simulation training have reported poorer training outcomes with VR compared to dry lab simulators[28, 118, 100, 181, 34, 116, 33]. Such differences have been attributed primarily to differences in haptic feedback between the simulator and dry-lab simulator. Proficiency in laparoscopy requires not only hand eye coordination but adaptation to the reduced haptic feedback from laparoscopic instruments. Tissue interactions and the corresponding haptic feedback have also been postulated to
play a role in training\cite{100}. Particularly in the early stages of training, realistic haptic feedback has been found to aid progression along the learning curve\cite{233}.

A number of studies have also reported lower participant rating scores of VR simulators, mirroring the inferior training outcomes\cite{28, 33, 34}. Feedback reports from studies comparing dry and VR modalities have also shown dry lab simulators to be more popular than VR with participants\cite{96, 143, 118, 225}. In all cases VR was criticised by trainees for its lack of realism. Even those VR systems that offered haptic feedback were considered unrealistic. These differences were also highlighted in a major review of VR and dry lab laparoscopic training\cite{164}. Interestingly these findings contrast with a survey of training directors which reported that VR was viewed as the second most important training modality; inferior only to operating room experience\cite{46}. In spite of these differences, a number of studies have shown that equivalent training outcomes are achievable with either dry lab or VR simulators\cite{163, 63, 8}. Even those studies discussed above reported only marginal differences between the modalities, with significant training effects seen with either dry lab or VR training.

Robotic surgery offers the unique challenge of operating in the complete absence of haptic feedback. Surgeons are forced to compensate by using visual cues such as the interaction with tissues and surgical materials as a surrogate for tactile feedback\cite{92}. This effect is seen even in novice surgeons\cite{92}. Significant advances have been made in robotic VR simulators which have been widely validated and offer numerous benefits for training\cite{19}. Simulators provide comprehensive objective assessment through performance metrics and instantaneous feedback. Specifically, in relation to robot surgery, the standalone simulators offer the great advantage of not requiring access to the Da Vinci console and thereby limiting training to times when the robot is not in clinical use. Another important consideration for robotic surgery is the high cost of consumables further limiting the application of dry lab simulation. Alongside these practical benefits of VR simulation for robotic
surgery, technological advances mean that the current graphical interfaces are able to offer very realistic imitations of the operative environment. A study, despite using a relatively weak methodology, has reported good participant reported ratings for the realism of robotic VR simulation programmes when compared to live surgery[108]. In support of this, and in contrast to the results of this study, a number of trials have shown equivalent outcomes for robotic VR and dry lab simulation training[239, 12]. Furthermore one study did reported higher participant satisfaction scores with VR over dry lab simulation[239]. Unlike the present study which utilised a complex suturing task as the final assessment, both trials used simpler assessment tasks. As a result, the marginal advantages of dry lab over VR simulation may not have become apparent in these limited results. For whilst the realism of VR simulators is good, the programmes still fail to fully match the almost limitless intricacies of real life. In more complex robotic tasks, a minor but key loss of visual detail can result in a significant restriction in operative ability particularly for inexperienced novice surgeons. This is reflected in the results of this study.

The results from this study do need to be considered in regards to a number of limitations. The sample size of this study was low and the dry-lab arm was underpowered however sensitivity analysis did demonstrate the results to be robust. Novice participants were selected to ensure as far as possible that all started at the same level. The disadvantage of being at the very beginning of the learning curve is that any educational intervention may lead to significant increases in performance limiting the discrepant effects of either training modality. Further analysis of more experienced surgeons is required to determine whether similar effects are seen. An alternative approach to overcome or at least limit this drawback of the study would have been to design a cross over study. However, the limited facilities in the two centres meant that not all participants had access to a VR simulator.
5.4 Discussion

5.4.1 Summary

This study has reported on the first comparative trial of dry-lab and VR simulation training. 43 robotic novices were recruited and trained at two centres in the UK and Italy. Participants were allocated to either VR or dry-lab simulation and completed a robotic simulation training programme. These results have shown that both dry lab and VR simulators are effective simulation training tools for basic robotic surgery. However, for more complex tasks such as robotic suturing, there is evidence to suggest that dry lab training offers advantages over VR training. Dry-lab training showed significantly greater improvements than VR simulation but only after the 3rd training session. It should be noted that both techniques were highly rated by all participants. The results of this study support the role of dry lab simulation even in the era of VR simulation. The true realism of dry lab simulation offers significant benefits and should form an important component of the robotic training curriculum.
Chapter 6

Cadaveric Wet lab Simulation Training

6.1 Introduction

Wet lab simulation in the form of cadaveric or live animal models represents the highest fidelity simulation available for training. Whilst virtual reality (VR) simulation continues to improve, as demonstrated in chapter 4, the accuracy of the tissue handling and anatomical detail that wet lab training offers remains unsurpassed. Wet lab simulation may refer to human cadaveric models, animal cadaveric models and live animal models. For the purposes of this study the definition will be limited to human cadaveric and live animal models. Whilst cadavers have been using in surgical training around the world for many centuries, in many areas it remained proscribed forcing the ardent surgeon and anatomist to obtain specimens by various nefarious, if not illegal methods. Indeed in the UK, it was not until 2006 that the Human Tissue Act 2004 came into force allowing surgical procedures to be performed on human cadavers[244]. Prior to this, anatomical dissection was permitted but any rehearsal of surgical techniques was strongly prohibited. This relaxation of the statute enabled the development of cadaveric training programmes in the UK for both surgeons and the many other healthcare professionals who continue to benefit from cadaveric simulation training[144].
Following enactment, there was a rapid growth in cadaveric simulation training programmes within the UK.

Although most anatomical dissection is performed on embalmed cadavers, the reduction in the tissue quality caused by the embalming process and the resultant rigidity largely precludes the use of embalmed cadavers for surgical training. Fresh frozen cadavers are used most commonly for surgical training. Specimens are immediately frozen to preserve tissue quality and then defrosted prior to the teaching event. This technique provides relatively realistic tissues but does come with a number of disadvantages. Principally specimens must be used within a 24-48 hrs following thawing otherwise tissue start to decompose. Nevertheless, it is frequently noted by participants that cadavers develop an unpleasant odour over the course of a training session which can be very off-putting. The very limited shelf life of cadavers also greatly limits their utility and necessitates cold storage facilities which all result in increased costs. It should be noted that for optimal preservation cool (4°C- 8°C) rather than cold storage is required and some authors recommend vacuum sealing to prevent mummification[177]. An alternative approach is Thiel embalming. This technique uses very low concentrations of formaldehyde alongside glycol and other salts[266]. This form of “soft” embalming preserves tissue texture and colour whilst avoiding the need for refrigeration or other special storage. Studies have demonstrated overall good fidelity of Thiel embalmed tissues compared to live or cadaveric tissue[80, 262, 67, 184]. Preservation of tissues with the Thiel technique also means that specimens can be reused (as long as they are not “over dissected”) and can used over longer periods of time. The drawbacks of this technique are firstly that costs of preparing the specimens are higher than either standard embalming or fresh frozen and some tissues such as brain, eyes are not amenable to the technique[200, 266].

The unique benefit of wet lab training, irrespective of the preservation technique is that it so closely mimics real life operating and as such offers an excellent platform
for training specific procedural skills[80]. Wet lab training is often considered the
gold standard for simulation training as it most closely models live patients[13, 26].
The accuracy of the specimens also allows their use in subspecialist fields to
train complex procedures and develop new techniques[102, 240]. However, for
basic skills training, whilst they can be trained used wet lab models, VR and
dry lab models are preferred. For basic training, it is argued that the training
environment and anatomical fidelity are less important and wet lab simulation of
whichever form remains a scarce and valuable resource[98, 223]. The value of both
cadaveric and live animal training has been shown for a number of modalities such
as laparoscopy, endourology and open surgery[128, 4]. Similar findings have been
found to a more limited extent in robotic surgery[26]. These results were reflected
in the conclusions of chapter 2 which demonstrated that wet lab training remains
the preferred technique for procedure specific training. To date the majority of
studies focus on basic and generic skill training and there is limited assessment of
the outcomes following cadaveric simulation particularly for robotic surgery[54]. In
particular there remains a lack of evidence to support the effectiveness of wet lab
simulation for advanced procedural training[83]. This study aims to evaluate the
effectiveness of a dedicated cadaveric simulation programme for advanced robotic
surgery training. Given the absence of established cadaveric robotics training
programmes available in the UK or Europe we developed and ran a new curriculum
programme for robotic cadaveric training and evaluated its effectiveness.

6.2 Methods

6.2.1 Curriculum

A panel of expert robotic surgeons and surgical educators designed the structure
of the course. Course content was primarily directed at advanced robotic training
for urological surgery. The course did provide generic advanced surgical skills training which allowed the inclusion of surgeons from other specialities. The training programme included knowledge based small group teaching, observation of live and as-live surgery (pre-recorded live surgery) and hands on training in the form of VR and cadaveric simulation training. The course was designed in accordance with the 3-step model established by Fitts and Posner [73]. Learning started with gaining an understanding of the task, followed by practice of the task. Further ongoing training would then lead to gaining the autonomous stage of learning.

Knowledge based training involved didactic teaching on set up and use of the surgical robotics, explanation of technical performances of the key surgical procedures covered in the training course (prostatectomy, cystectomy with lymph node dissection and radical/partial nephrectomy) as well as expert advice (such as “tips and tricks”, avoiding and resolving complications). Non-technical skills training specific to robotic surgery was provided during the hands-on training. Didactic teaching was organised as a series of short presentations with a focus on group discussion. Sessions aimed to establish a baseline understanding of the core concepts in robotic surgery together with an explanation of the necessary technical and non-technical skills required for the successful completion of the various procedures. Common pitfalls and points of failure were identified in the “tips and tricks” sessions; recognition of frequently occurring error is an important aspect of training[207]. This knowledge-based training was supported by observation of live and as live surgery of the procedures discussed in the preceding sessions. Faculty were selected on the basis of their sub-specialist expertise in robotic surgery. Experts from other fields such as anaesthetics and non-technical skills training were also recruited to provide the necessary expertise (see Appendix C.4, p. 251.
Hands on training consisted of virtual and cadaveric simulation training with a focus on cadaveric training. Virtual reality simulation was used for warm up sessions and enabled participants to re-familiarise themselves with the surgical robot. Cadaveric simulation training was conducted in pairs of participants under the proctorship of an expert robotic surgeon. During each session participants would complete a section of the procedure being trained. This allowed maximisation of both training time and efficient usage of the cadavers helping to reduce course costs. Each cadaver was used to teach pelvic (radical prostatectomy; radical cystectomy; extended lymph node dissection) and renal procedures (partial nephrectomy; radical nephrectomy). During each session participants were assessed using GEARs and NOTSS. The video feed from the endoscopic cameras was recorded for post course blinded assessment.

Courses were advertised online and through medical societies. Attendance was open to senior trainees and consultant surgeons. No minimum levels of experience were set. Course fees were strictly controlled to enable maximal participation and charitable funding was organised for UK based trainees for certain courses.

### 6.2.2 Hands on Cadaveric Simulation Training

Cadaveric training took place in dedicated anatomy laboratories. Four to five fresh frozen cadaveric torsos were used for each training course. Cadavers were thawed at room temperature three days prior to use. Each cadaver was prepared in advance to minimise disruption to training. The training environment was made as realistic as possible with the use props such as drapes, theatre trolley and instruments. During each training session the mentor would guide pairs of participants through the procedural steps. Preceding lectures were designed to complement hands on training. Prior to training and between sessions participants
were given the opportunity to undertake further VR simulation training for basic skills training.

### 6.2.3 Evaluation of the Training Course

The training course was evaluated through a combination of quantitative and qualitative outcome measures. All participants completed a questionnaire before the course detailing their demographic information, surgical experience and also outlined the aims and structure of the course. During the course, participants’ technical skills were objectively assessed using GEARS. Video recordings of participant performances, taken from the robotic endoscopic camera, were blindly assessed by a rater experienced in robotic skills assessment using GEARS. Technical performances during the first 15min of the first training session were regarded as reflective of the baseline, pre-training ability of the participant. The final 15min of the last training session on the second day was used as a proxy for post training technical ability. Following the course, participants completed a further questionnaire evaluating their experience of the course together with reflections on their progression during the course and impact of the course.

### 6.2.4 Statistical Analysis

Given the non-parametric data, linear mixed models were used to evaluate pre and post course GEARS score across the courses controlling for dry lab simulation and robotic experience. On examination of the change in GEARS score pre and post course, residual errors were found to be normal (see Appendix C.1, C.2). Further analysis of the relationship between the technical progression during the course and previous dry lab experience and robotic surgical experience was also undertaken using linear mixed models. Mann Whitney analysis was used to analyse training effects of the course. Statistical analysis was undertaken using
6.3 Results

6.3.1 Baseline Participant Characteristics and Technical Ability

The results of five courses run between 2015 and 2017 were analysed. In total they provided training for 50 participants. Demographic data was available for 43 participants (86%) (Table 6.1, p. 132).
### Table 6.1 Demographic Details and Surgical Experience

<table>
<thead>
<tr>
<th></th>
<th>Newcastle Freman 2015</th>
<th>Newcastle Freman 2016</th>
<th>Guy’s and St Thomas’ 2015</th>
<th>Guy’s and St Thomas’ 2016</th>
<th>Guy’s and St Thomas’ 2017</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Participants</td>
<td>16</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>n/a</td>
</tr>
<tr>
<td>Mean Age</td>
<td>45.8</td>
<td>37.3</td>
<td>45.1</td>
<td>46.8</td>
<td>39.2</td>
<td>0.007</td>
</tr>
<tr>
<td>Male: Female</td>
<td>7:1</td>
<td>7:1</td>
<td>8:0</td>
<td>9:0</td>
<td>8:0</td>
<td>0.5</td>
</tr>
<tr>
<td>Consultant: Trainee Ratio</td>
<td>16:0</td>
<td>0:8</td>
<td>7:1</td>
<td>3:2</td>
<td>3:5</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>Experience of Consultants (mean years)</td>
<td>9.5</td>
<td>n/a</td>
<td>9.9</td>
<td>8.5</td>
<td>15</td>
<td>0.95</td>
</tr>
<tr>
<td>Clinical Experience of Trainees (mean years)</td>
<td>n/a</td>
<td>7.1</td>
<td>8.0</td>
<td>7.25</td>
<td>7.2</td>
<td>0.84</td>
</tr>
<tr>
<td>No. of Participants who have assisted in Robotic Surgery</td>
<td>5 (31.3%)</td>
<td>7 (87.5%)</td>
<td>4 (50.0%)</td>
<td>4 (40.0%)</td>
<td>6 (75.0%)</td>
<td>0.005</td>
</tr>
<tr>
<td>No. of Participants who have performed Robotic Surgery with Guidance or Independently</td>
<td>3 (18.8%)</td>
<td>4 (50.0%)</td>
<td>1.0 (12.5%)</td>
<td>2 (20.0%)</td>
<td>5 (62.5%)</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean No. of Cases Assisted ± SD</td>
<td>14.5 ± 49.7</td>
<td>95.3 ± 54.4</td>
<td>140.7 ± 344.1</td>
<td>n/a</td>
<td>102.2 ± 131.0</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean No. of Cases Performed with Guidance ± SD</td>
<td>2.3 ± 6.7</td>
<td>10.8 ± 24.2</td>
<td>0.8 ± 2.12</td>
<td>n/a</td>
<td>14.2 ± 14.0</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>No. of Participants with VR Simulation Experience</td>
<td>7 (43.8%)</td>
<td>7 (87.5%)</td>
<td>8 (80.0%)</td>
<td>5 (50.0%)</td>
<td>4 (50.0%)</td>
<td>0.2</td>
</tr>
<tr>
<td>No. of Participants with Dry Lab Simulation Experience</td>
<td>7 (43.8%)</td>
<td>3 (37.5%)</td>
<td>8 (80.0%)</td>
<td>5 (50.0%)</td>
<td>4 (50.0%)</td>
<td>0.4</td>
</tr>
<tr>
<td>No. of Participants with Wet Lab Simulation Experience</td>
<td>0</td>
<td>2 (25.0%)</td>
<td>2 (25.0%)</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>No. of Participants having Completed a Robotic Fellowship</td>
<td>1 (6.2%)</td>
<td>2 (25.0%)</td>
<td>0</td>
<td>2 (20.0%)</td>
<td>1 (12.5%)</td>
<td>0.6</td>
</tr>
<tr>
<td>Virtual Reality Simulation Training (mean hours ± SD)</td>
<td>3 ± 7.4</td>
<td>22.9 ± 23.5</td>
<td>7.9 ± 12.0</td>
<td>2.6 ± 3.3</td>
<td>7.8 ± 9.0</td>
<td>0.05</td>
</tr>
<tr>
<td>Dry lab simulation (mean hours ± SD)</td>
<td>0.5 ± 1.4</td>
<td>1.9 ± 3.0</td>
<td>3.8 ± 5.9</td>
<td>0.9 ± 1.4</td>
<td>2.0 ± 4.2</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 6.1 (continued)
### Table 6.1 Demographic Details and Surgical Experience

<table>
<thead>
<tr>
<th></th>
<th>Newcastle Freeman 2015</th>
<th>Newcastle Freeman 2016</th>
<th>Guy's and St Thomas' 2015</th>
<th>Guy's and St Thomas' 2016</th>
<th>Guy's and St Thomas' 2017</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Participants trained in Laparoscopic Surgery</td>
<td>9 (56.6%)</td>
<td>5 (62.5%)</td>
<td>8 (88.9%)</td>
<td>8 (80.0%)</td>
<td>4 (50%)</td>
<td>0.2</td>
</tr>
<tr>
<td>Mean Years of Laparoscopic Experience ± SD</td>
<td>3.2 ± 3.8</td>
<td>0.8 ± 0.4</td>
<td>8.3 ± 6.1</td>
<td>6.5 ± 6.0</td>
<td>4.5 ± 2.6</td>
<td>0.1</td>
</tr>
</tbody>
</table>

The End
Alterations in funding for the various training courses resulted in significant variations in participant composition over the five courses. Participants remained predominantly male and both trainees and consultants had similar levels of respective experience. Whilst there were significant variations in the robotic experience (both as assistant and operating surgeon), levels of simulation experience were equally low across the groups. Laparoscopic experience was also relatively equal.

Possible factors influencing baseline technical ability were investigated. Initially univariate regression analysis of all possible factors collected in pretraining questionnaires was performed. Significant interactions were seen between baselines GEARS scores and robotic surgical experience, VR simulation experience and dry lab simulation experience. Notably assisting experience, age, laparoscopic experience, wet lab simulation experience and previous robotic fellowships had no relationship with baseline GEARS score. Multivariate analysis was performed incorporating the following significant factors; robotic surgical experience, VR simulation experience and dry lab simulation experience. It was found that robotic surgical experience ($\beta = 0.53, p = 0.01$) and dry lab simulation experience ($\beta = 0.50, p = 0.05$) were significant predictors for baseline GEARS score. VR simulation experience was not a significant predictor ($\beta = 0.09, p = 0.96$). Overall the model account for a significant proportion of variance in baseline score, adjusted $R^2 = 0.48, p < 0.005$.

6.3.2 Assessment of Technical Performance

Overall comparison of pre and post wet lab training scores demonstrated a significant improvement in GEARS score (Wilcoxon signed rank test $Z = -4.2, p < 0.005$). It was noted that there were significant variations in both baseline and final GEARS score between the different courses. Theses significant differences in the baseline and end of training GEARS scores between the different courses
6.3 Results

persisted despite controlling for robotic surgical and dry lab simulation experience (p = 0.03 and p = 0.004 respectively). When the absolute improvement in GEARS score was analysed (post course minus pre-course GEARS score) no difference was seen between courses, p=0.33.

A linear mixed model was used to assess pre-post GEARS score controlling for dry lab simulation and robotic surgical experience. As shown with the Wilcoxon signed rank test, there was a significant increase in mean GEARS score over the course (mean increase GEARS score 2.9 95% CI 1.9-3.8 p <0.001). Prior dry lab simulation or robotic surgical experience resulted in a loss in this difference primarily due to increases in baseline performance (Appendix C.3). To explore these differences further, dry lab and robotic experience were converted to categorical variables (dry lab simulation was divided in no experience/ 1-5 hours/ > 5 hours; robotic experience divided into no experience/ 1-25 cases/ > 25 cases. Graphical analysis (Figure 6.1, p. 135; Figure 6.2, p. 136) of training stratified by dry lab simulation and robotic surgical experience demonstrated increases in GEARS following training for all levels of experience. Participants with no or little (less than 5 hours/25 cases) of dry lab or robotic surgical experience respectively scored significantly lower prior to training. In contrast whilst absolute post training scores were lower for participants with no or little dry lab or robotic experience compared to participants with greater experience these variations were not significant. Further comparison of categorical experience levels against change in GEARS score was performed. There were no significant differences in the change in GEARS score across the 3 strata for either dry lab or robotic experience (Table 6.2, p. 135).
6.3 Results

Table 6.2 Change in Mean GEARS Score Stratified by Dry Lab and Robotic Surgical Experience

<table>
<thead>
<tr>
<th>Dry Lab Simulation Experience</th>
<th>No experience</th>
<th>1-5 hours</th>
<th>&gt;5 hours</th>
<th>Robotic Surgical Experience</th>
<th>No experience</th>
<th>1-25 cases</th>
<th>&gt;25 cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.0</td>
<td>1.6</td>
<td>2.8</td>
<td></td>
<td>2.7</td>
<td>2.6</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2</td>
<td></td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6.1 Change in GEARS Score Stratified by Dry Lab Experience
6.3 Results

136

Fig. 6.2 Change in GEARS Score Stratified by Robotic Surgical Experience

6.3.3 Post Course Evaluation

Participants completed a post course questionnaire addressing their opinions on the cadaveric models, the simulation training programme and the role of cadaveric simulation in surgical training. Overall participants scored both the realism of the cadavers and the procedures very highly with mean scores ranging from 3.9/5 to 4.5/5 (Figure 6.3, p. 137).
Fig. 6.3 Opinions on the Fidelity of Cadaveric Simulation Training
When questioned on the effectiveness of cadaveric simulation training, participants similarly recognised its benefit. All participants agreed strongly that cadaveric simulation was a good training tool for robotic surgery and 81% agreed strongly that they felt more confident in performing robotic surgery having completed the cadaveric training programme. Furthermore 93% of respondents agreed strongly that cadaveric simulation would be effective for robotic skills assessment and 75% agreed that cadaveric training was equivalent to operative experience. Overall participants scored their enjoyment of the course 4.6/5. The course price was scored 3.3/5 (5 = very expensive 1 = very cheap) whilst the duration of the course also scored well (2.9/5 (5 = too long 1= too short).

6.4 Discussion

This study reports the first validation of cadaveric simulation training for robotic surgery. The results have shown both quantitative evidence that cadaveric training results in improved technical performance as well as qualitative evidence that the course offers effective, high fidelity training that is highlight regarded by participants.

Assessment of the relationship between participant related factors and pre-course surgical ability demonstrated a significant relationship only for previous robotic surgical experience and dry lab simulation experience. Interestingly VR simulation experience, prior laparoscopic or robotic assisting experience did not influence baseline robotic surgical ability in this study. Objective assessment of technical performances before and after training show significant increases in GEARS score when controlled for robotic surgical experience and dry lab training. Further analysis of participant performances in the context of these significant variables proved that the cadaveric training did lead to increases in technical ability as shown by higher GEARS score. Interestingly variations in these factors
influenced pre and post training performance although equal rates of improvement were seen at all levels of experience. This linear relationship between prior robotic experience and course performance in particular provides evidence for the external validity of cadaveric training.

Qualitative evaluation of the effectiveness of the training course and the fidelity also demonstrated positive results. The majority of participants scored the cadaveric models and the training procedures highly for realism and importantly the ability to dissect and identify tissues. Similarly, all the procedures were also rated very realistic by the participants. These data provide substantial evidence for the content validity of cadaveric models for robotic training.

The feasibility of using cadaveric models for simulation training has been demonstrated previously in other specialities. Cadavers have been shown to be effective training tools for basic skills in laparoscopy[133, 223]. Studies have investigated the role of cadaveric training for procedural skills training. Reflecting the results of this study, they have shown it to be an effective training modality, superior to VR simulation[80, 134, 222]. These studies highlighted the tissue quality and tissue planes as important factors missing in even high-fidelity VR simulation. The qualitative feedback from this study supports these findings. These qualities of the cadaver help participants develop the necessary knowledge of surgical anatomy, technique and use of surgical instruments. It is often reported that a major drawback of cadaveric training is the lack of bleeding however this can be a valuable attribute. Managing bleeding is a critical aspect of surgery however for training the absence of bleeding can be beneficial, allowing trainees time to explore tissues clearly[121]. Importantly cadaveric training has been also shown to have high satisfaction scores and is generally preferred to VR simulation training[133, 134, 222]. Although beyond the scope of this study, limited evidence has been reported to support the consequences of simulation training. Two studies
have reported limited evidence to show improvement in clinical skills following cadaveric training[81].

The other major drawback to cadaveric training is the cost and facilities required to provide training. As discussed above a cadaveric laboratory is required to prepare and store cadavers irrespective of the embalming technique. The associated costs of acquiring and preparing the specimens for training are high. Cadaver costs for the courses I organised for this study were between £2000-£2500, in line with the literature[26, 133]. All attempts were made to reduce costs such as carefully planning training to maximise use of the cadavers. Despite this, for this study training costs remained high with total costs of £18,000-£20,000 per course. Procurement of cadavers can also be problematic, and is reliant on the public bequeathing their bodies to science. In the UK, the supply is limited and most be shared amongst medical schools and other bodies undertaking cadaveric research and training[81]. It can be even more difficult in countries where cultural practices value the integrity of the body even after death. Some authors have reported more entrepreneurial approaches such as unclaimed bodies in hospital mortuaries[235]. In this study, as a result of the limited number of specimens within the UK, I was forced to import specimens from the US further increasing costs. Another limitation to cadaveric training are the ethical and safety considerations. It was not reported by participants in this study, but some may be reluctant to use human cadavers to practice their surgical skills. A number of studies have reported concerns from participants over the safety of practicing on cadavers. Sharma et al. [223] reported that 3/45 participants voiced concerns over contracting infections from the cadavers and the hazards of inhaling gas from insufflation. No participants in this study had such concerns but they do need to be taken into consideration when developing training curricula.

This study involved the establishment of a novel cadaveric training programme for robotic surgery, not previously undertaken in the UK. The course was carefully
developed following extensive review of training practices as discussed in Chapter 2. This study has demonstrated the feasibility of delivering cadaveric robotic training courses and initial results have shown the validity of this training in a number of areas. The limitations, however, must be considered. As a result of the focus on delivering high quality training, further validity assessment of cadaveric training is required. I have demonstrated some evidence to support the educational impact of training, but more objective assessment of performance outcomes following training are required.

6.4.1 Summary

As part of this study, a novel cadaveric robotic training programme was developed. The course consisted of a didactic lectures, observation of live and as-live surgery and hands on VR and wet lab robotic surgical training. The course was directed towards senior surgical trainees and consultant surgeons with an interest in developing a robotic surgical practice. Five courses were delivered between 2015 and 2017. Qualitative and quantitative data from 50 participants was collected. Analysis of the data demonstrated evidence to support the validation of cadaveric robotic simulation training. Qualitative data from participant questionnaires endorsed the feasibility and acceptability of delivering robotic cadaveric simulation training. Developed using an evidence-based approach, the training programme was well received by participants who acknowledge the beneficial training outcomes at an affordable price. Quantitative data provided initial results supporting the content and external validity of cadaveric training. Further work needs to be undertaken to further establish the full construct validity of cadaveric training for robotic training but these initial results strongly support its inclusion on the robotic curriculum.
Chapter 7

Wet Lab Training: Live Animal

7.1 Introduction

Like cadavers, animals are one of the oldest "simulation" tools used by surgeons to practice and develop their skills. Yet with the development of virtual reality and synthetic models, wet lab simulation using live animals has become a relatively uncommon training modality. Animal models offer various benefits but there are numerous moral, ethical and regulatory limitations that must be concerned. Nonetheless animal models offer some of the most realistic training models available to surgeons and their teams. As stated previously, simulation training must engage with the complexity and unruliness of real life[126]. The main advantages of animal models are both the anatomical similarity and more importantly the verisimilitude of the tissues. A wide range of skills such as port placement to provide optimal exposure, cutting, dissection, coagulation and suturing may be practised. Furthermore, animals are relatively unique in their ability to model bleeding and other intraoperative complications: "complication training remains a major application of live animal training[97]. One of the greatest benefits of live wet lab training is the interaction and reaction of the training "models. Despite the advances in other types of simulation, physiological responses such as bleeding, if
present, remain crude and appreciably different to real life. As a result, the use of live animals for surgical training continues to play an important role in simulation training. Reductions in working hours and consequently clinical training have had major impact on training for such situations, with a lack of exposure to the major but rare complications. Yet despite the accurate anatomy and realistic physiological responses, animal training still does not fully recreate the stress of managing complications in the operating room (OR)[54].

Wet lab simulation requires dedicated facilities and personnel so that training can be conducted in accordance with the various ethical and regulatory requirements. The use of animals for training, as for all animal experimentation, must be carefully considered. The 3 Rs, initially described in 1959, still guide regulations in Europe[208]. These stand for replacement, reduction and refinement and are recognised as key strategies for maintaining humane animal experimentation. Replacement refers to the aim of trying to avoid the use of live animals. The concept of reduction is to implement strategies that enable the use of as few animals as possible. The final component, refinement, refers to the modification or adaption of animal care methods to minimise as far as possible any pain, suffering or stress that the animals may experience. In the past “chronic animal models” were used whereby the animals would undergo multiple operations over the course of a number of weeks interspersed with periods of recovery[265]. The resultant trauma and stress on animals have thankfully lead to this practice being abandoned. Similarly previous techniques whereby animals undergo surgery to create a create pathology that then is used to train such as tying the common bile duct in pigs is also now prohibited[40].

The principle arguments against the use of live animals for surgical training encompass the ethical and moral objectives to harming animals. If suitable alternatives exist, it is difficult to justify the use of live animals for simulation purposes. The 2002 statement from the American College of Surgeons outlines
a similar position that “wherever feasible, alternatives to the use of live animals should be developed and employed”. The college does go on to say that “now and in the foreseeable future it is not possible to completely replace the use of animals . . . . [in] education and teaching”. Hence, until simulators are able to realistically model the unruly human body accurately, a role for live animal training will persist.

For laparoscopic and robotic training, porcine models are most commonly used given the anatomical similarities to humans. Use of a range of animals has been reported including dogs, cats, sheep and rabbits[265, 49, 40, 97]. Chicken models have also been used extensively, and in imaginative ways such as using the crop filled with water to simulate the renal pelvis[199]. Dog models have been used for vascular, urological and general surgical training, rabbits and rats for microsurgical training and cats for paediatric surgery[265]. Variations in anatomy such as stomach and bowels, Calot’s triangle, CBD and less distinct prostate glands need to be considered[40, 80].

In contrast to much of the rest of the world, the use of live animals for surgical training is still proscribed in Britain. Live animal models are more widely used for training in the US and across Europe. Due to the necessary infrastructure and facilities to meet the ethical and regulatory standards for caring for the animals, the availability of animal training remains limited. Given the similarities to human anatomy, pigs are most commonly used. Some anatomical variations do however exist limiting the training that can be undertaken with porcine models[120]. For example, pigs have no retroperitoneal fat and the kidney is directly adherent to the posterior abdominal wall. The spleen is also not as wide and much more mobile than in humans again altering the intra-abdominal space. Cadaveric animal tissue has is more widely used given its easy availability and the lack of regulatory constraints or specialist facilities. In comparison to live animals which provide both close anatomical fidelity and live tissue or human cadavers, cadaveric animal
tissue remains significantly inferior. As a result, this study was limited to live animal models.

This study aimed to evaluate the role of live animal simulation for robotic surgical training through a qualitative and quantitative survey of participants at a live porcine robotic training course.

7.2 Methods

7.2.1 The Live Porcine Robotic Simulation Course

The porcine robotic training programme consists of a two day hands-on training course held at the Minimal Invasiv Udviklings Center, Aalborg University Hospital, Aalborg, Denmark. The course programme was developed with a focus on hands on training and was aimed at surgeons with at least an intermediate level of skill in robotic surgery. The programme includes a mixture of short lectures on technical performance interspersed with sessions of robotic training which comprises the bulk of the course. Participants are divided into pairs alternating between working on the console and assisting. Proctoring is provided by expert robotic surgeon and an experienced bedside assistant is present (Figure 7.1, p. 146).

In accordance with the 3Rs, the course is held in the biomedical research laboratory, a purpose-built unit with the necessary facilities and equipment to provide safe and ethical care for the animals. Alongside surgical training, the laboratory also conducts regular scientific studies involving animals such as domestic pigs, mini-pigs, rabbits, guinea pigs, rats and mice. As a result, laboratory staff have extensive experience in caring for animals. Official approval is provided by the Danish Animal Experiments Inspectorate. Throughout each course an experienced veterinarian manages the animals, providing anaesthesia, physiological support and eventual euthanasia.
Fig. 7.1 Example of Porcine Robotic Surgical Training at Minimal Invasiv Udviklings Center, Aalborg University Hospital
7.2.2 Development of the Participant Survey

An online evaluation for participants who had completed the course over the preceding 10 years was conducted. A questionnaire was developed in cooperation with surgical training experts (Appendix D, p. 258). It consisted of 18 questions focussed on the participants prior robotic surgical experience, self-assessed level of robotic skill before and after completing the training programme, and a comparative assessment of the training course. Respondents were asked to self-assess their surgical ability and confidence out of 100. 5-point Likert scales were used to rate the educational benefits of the individual course components. Participants were contacted directly by email and invited to complete the survey using an internet link. Two reminder emails were sent to all participants who did not reply.

7.2.3 Statistical Analysis

Data analysis and graphs were plotted using GraphPad Prism version 8.0 for macOS, GraphPad Software, La Jolla California US Results. Self assessment scores were compared using Mann Whitney U test and correlations between final assessment scores were test using multiple linear regression analysis.

7.3 Results

174 participants attended the course between 2009 and 2014. Contact details were available for 140 people all of whom were invited to take part in the survey. 39 (28%) of contacted participants responded and completed the full survey. There was a range in experience levels of participants from junior surgical trainees to consultant surgeons (Figure 7.2, p. 148).
Fig. 7.2 Aalborg Participants’ Training Experience
72% (n=28) of respondents came from the UK with the remainder from Denmark (23%, n=9), India (n=1) and Malaysia (n=1) surgeons. Self-rated robotic technical skill and confidence were both low before the course (mean 3.5 ± 6.8 and 3.5 ± 7.4 respectively) (Figure 7.3, p. 150).
Fig. 7.3 Pre-Course Self Rated Robotic Skill
Similarly, the overwhelming majority of participants did not have any console experience prior to the course, mean of $8.9 \pm 22.1$ cases with console experience (Figure 7.4, p. 152).
7.3 Results

Fig. 7.4 Pre-Course Robotic Console Experience
74% of participants (n=29) had undertaken some form of robotic training before the porcine robotic course. 41% (n=16) used just one form of training and a further 8 participants used 2 training modalities (21%) (Figure 7.5, p. 153).
7.3 Results

7.3.1 Outcomes of Porcine Robotic Simulation Training

Participants were questioned on the perceived usefulness, effectiveness and realism of the training course. With regards to the individual components of the training course, responses were overwhelmingly favourable (Figure 7.6, p. 155). Across all procedures taught on the course, the majority of participants found them very useful. The most useful were port placement and docking, basic robotic skills training and repair of a bladder injury.

Participants were also asked to rate their own robotic surgical ability and their confidence in performing robotic surgery following the training course. 72% of participants (n=28) reported an increase in their perceived ability and 79% (n=31) reported an increase in confidence in their robotic skill. Overall there was a statistically significant increase in both self-rate technical ability and confidence following the course (Figure 7.7, p. 156).

The relationship between self-assessed technical ability/ confidence post training and pre-course technical ability/ confidence and robotic surgical experience was modelled using multiple regression analysis. For self-assessed technical ability, there was no relationship to pre-course self-assessed ability (p=0.06) but robotic experience did significantly contribute to post course self-assessed technical ability (p=0.05). In contrast post confidence was related to pre course confidence (p=0.005) but not robotic experience (p=0.24).

The final questions in the survey regarded the participants further robotic training and practice after the course. The majority of participants took part in further training after completing the Aalborg Programme (n=24, 71%). Most tried 1 modality of training (n=12) with 8 and 4 participants using two and three other modalities of training respectively. Most participants went on to undertake operating room training (Figure fig:postcourseoutcomes, p. 158). VR Training, Dry lab training and further live animal training was also used. No participants
Fig. 7.6 Participant’s Perceived Usefulness of Procedural Training and Overall Realism of Porcine Training
7.3 Results

Fig. 7.7 Change in Participants’ Self Rated Ability and Confidence
underwent cadaveric training. Following the course, the majority of participants continued to either train or perform robotic surgery whilst 15 (38%) no longer undertook robotic surgery.
7.3 Results

Fig. 7.8 Post Course Training and Robotic Experience
7.4 Discussion

A cross-sectional survey was used to analyse both the population of surgeons undertaking live animal robotic surgical training and the effectiveness of the training programme. The study has outlined a number of important insights into porcine robotic training. Regarding participants, despite the course providing relatively advanced robotic skills training, itself a sub-specialist surgical field, the majority of participants were in the early stages of their residency training. This was further reflected in low pre-course self reported robotic skill and robotic surgical experience. Trainees also had undertaken limited amounts robotic training prior to completing the course. Over 25% of participants (n=10) had no robotic simulation experience. 41% (n=16) had used only one simulation training modality of which 6 participants had used only e-learning and undertaken no practical training. In total 16 participants (41%) had not previously undertaken any hands-on training in robotic surgery and had a median console experience of 0 robotic cases.

Despite low levels of prior experience, participants overwhelmingly scored the procedural training offered in the course as very useful. Interestingly basic skills training was best scored with 77% (n=30) of participants scoring it highly useful. Basic skills training was the most performed activity of the courses, completed by 38/39 participants. Of the procedural training components, repair of bladder injury, pelvic lymph node dissection and ureteric reimplantation were considered to be the most useful. Nephrectomy, salpingectomy and hysterectomy were scored the lowest. The low scores of the latter two procedures was likely due to majority of participants being urologists and experience in these procedures offers little utility. It should be noted that whilst a structured training is provided using a combination of didactic lectures and proctoring, hands-on training is directed
towards the individual requirements of the participants. Hence these results do offer a reflection of the training needs and experience of the participants.

The realism of porcine training was recognised by the majority of participants. This further reflected in the significant improvements seen in self rated technical ability and confidence following the course. Yet in spite of the above and whilst the majority of participants (71%, n=24) undertook further robotic simulation training, only 49% of participants continued to train or practice in robotic surgery.

As discussed in Chapter 6, p. 124, wet lab training faces considerable barriers. The principle barriers to live animal simulation are the costs and capital requirements necessary to support an animal laboratory. As discussed above full veterinary support is indispensable. From a practical point of view, an appropriate simulation centre is also important. Often simulation laboratories are relegated to a dark corner of the hospital, but for animal training well ventilated and spacious premises are required[180]. As a result, it is important to establish the unique benefits of using live animals over other forms of training. The advantages of responsive tissues and the ability to manage complications are not relevant to all trainees. For initial basic training, the emphasis lies on gaining psychomotor skills that can be provided through VR and dry lab bench-top models. It is not surprising that studies of live animal training frequently report high satisfaction scores; the ability to practise on a live animal is clearly attractive[225, 158]. Shetty et al. [225] postulated that these differences were due to the more rudimentary dry and VR models being less stimulating. It is therefore important to note the lower satisfaction scores reported in this study and the nephrectomy training in particular. These results are likely to be due to anatomical variations in swine. The absence of retroperitoneal or perinephric fat in pigs means that nephrectomy is substantially easier to perform than in humans. Katz et al. [120] compared cadaveric to live animal urological laparoscopy training (renal and bladder procedures) and found reported that human cadavers were preferred over live animals. Live
animal training has been compared to bench top simulation RCTs previously albeit not for laparoscopic or robotic surgery. Results were limited by the use of inexperienced participants but 3 studies of emergency bedside and endovascular procedures showed no significant differences difference[21, 93, 114]. In contrast a study of cardiovascular surgeons with some prior robotic experience (mean 1.7hr of robotic training experience) showed that wet lab training resulted in the greatest improvements in skill compared to dry lab or VR training[246].

The conclusions that can be drawn from this study are limited by the methodology. A wide and relatively large cohort of participants were included in the survey in comparison to similar studies[174, 231]. However the low response rate means selection bias needs to be considered. It can be hypothesised that those participants who did not engage with further training would be less likely to respond to this survey. Participant satisfaction rates must also be considered in light of potential bias. The survey was designed to avoid leading questions and required a variety of response to prevent acquiescent responses. The high rate of responses to blank box questions supports the participation of respondents. Clearly self-reported outcomes are not equitable to objective assessments of training outcomes but do provide useful insights in the perceived usefulness and acceptability of training. Finally as the survey included participants from a number of years, recall bias needs also to be considered.

7.4.1 Summary

A cross-sectional survey of 39 participants who underwent live porcine robotic training was performed. The results demonstrate live porcine robotic surgical training to a widely valued, acceptable and feasible form of advanced robotic skills training. Despite the relative inexperience of participants prior to undertaking the course, significant improvements in self assessed outcomes were reported. This
was mirrored by high rates of further training and robotic surgical experience undertaken by the trainees. The results of this survey support the use of live porcine training for robotic surgery. Limitations principally due to the cost and infrastructure required to provide means that careful consideration is required for its implementation in robotic training programmes. Training is best delivered to intermediate and advanced robotic training to ensure the unique benefits of live animal training are realised.
Chapter 8

Development and Validation of a Tool for Non-Technical Skills Evaluation in Robotic Surgery-
The ICARS System

8.1 Introduction

Establishing routine non-technical skill (NTS) training for surgery has been relatively slow in comparison to other high-risk industries such as the aviation. Nevertheless there is now widespread recognition of the importance of NTS alongside technical skill training replacing previous misconceptions that NTS are inherent behaviours that do not require formal training. As a result in the last decade there has been the development of increasingly sophisticated simulation programmes focussed towards NTS training[270]. In spite of this, the current academic literature continues to focus on arguments for the introduction of NTS training rather than further discussion on the development and validation of effective training tools.

The growing recognition of the role of NTS is clearly reflected in the quality and quantity of the behavioural training tools now available[264]. As discussed in earlier chapters, effective and accurate evaluation of simulation trainees is essential
for a number of reasons. It allows training to be monitored to ensure it remains effective, it provides goals to motivate participants and feedback can be used to help guide their further training. In particular for non technical skills training, feedback is important. The structured debrief is one of the main opportunities for learning and therefore occupies a central role in an NTS course. A debrief is a guided discussion of the scenario undertaken in an environment that fosters discussion. The performance of the participant or participants who have completed the scenario is analysed to identify strengths and weaknesses and strategies for improvements in the future. Behavioural rating systems play an important role in debriefing, providing a structured approach both to assessment of NTS and the debrief.

Provision of safe and reliable patient has always been of paramount importance in surgery, but this has been matched in recent years by large expansions in the regulatory monitoring of outcomes. Such scrutiny has helped to highlight that that technical competency alone does not guarantee success[5]. Recognition of the vital importance of NTS has resulted in the development of various NTS behavioural rating systems. Separate systems have been developed for assessment of the entire surgical team[245, 154, 206] as well as individual team members such as surgeons[269, 152], anaesthetists[74] and scrub practitioners[155]. To be effective such rating systems must accurately capture relevant NTS behaviours and notably there is significant overlap in the NTS that the various tools identify and measure. These similarities help to demonstrate the generalisability of NTS across both surgical specialities and surgical teams. Established systems such as Non-Technical Skills for Surgeons (NOTSS) taxonomy have also been applied to a variety of surgical specialties and environments outside the operating such as critical care. Alongside a validated behavioural rating system, trained faculty with experience of NTS assessment are also important. The participants themselves must also be motivated and understanding of the learning process.
Robotic surgery has expanded rapidly in recent years despite high costs limiting its use to specialist centres. Given the limited availability of robotic systems, training remains challenging and various simulation-based training curricula have been developed[72]. Robotic surgery demands significant adaptations to the standard operating room (OR) environment including team interaction. As a result, proficiency in robotic surgery requires specialist training in both technical and non-technical skills. Despite this, only recently have robotic surgical curricula begun to incorporate NTS[44, 161]; no behavioural markers systems have yet been developed for robotic surgery[264]. Whilst generic systems such as NOTSS can be used in robotic surgery, the unique environment presents its own NTS challenges. Previous studies have shown that in highly specialist environments generic assessment tools may not be suitable[104]. This study aims to develop the first behavioural rating system specifically for surgeons NTS during robotic surgery and provide initial validation evidence to demonstrate its applicability to this unique operating environment.

8.2 Materials and Methods

8.2.1 Development of the Interpersonal and Cognitive Assessment for Robotic Surgery (ICARS) Behavioural Rating System

This prospective study was conducted in collaboration with the Urological Department, Guy’s Hospital, the Vattikuti Robotic Surgery Training Centre and the MRC Centre for Transplantation, King’s College London. Approval was granted by the local research ethics committee.

ICARS was developed using a similar core framework to those described by previous role specific behavioural rating systems[269, 74, 155]. Development of
ICARS was performed in three stages. In the first stage, a full taxonomy of NTS behaviours relevant to surgeons performing robotic surgery was compiled. Secondly the key NTS behaviours in robotic surgery were identified using a modified Delphi process involving a panel of expert surgeons. In the final stage, a behavioural marker system was designed, incorporating the key NTS skills for robotic surgery into a practical checklist. The checklist was reviewed by the consensus panel for final approval.

To create a comprehensive catalogue of non-technical skills relevant to robotic surgery, 15 hours of live robotic surgery were observed by two study investigators trained in surgical NTS. All potential behaviours and skills relevant to non-technical skills assessment observed by the two study investigators were recorded and collated. Throughout, interviews were conducted with the operating surgeons following each case to gain further insight into potentially relevant NTS behaviours. These observations were then correlated with clinical observations. This provided a comprehensive record of non-technical skills relevant to robotic surgery. The final list was decided by agreement of the researchers before being grouped and sub-categorised into a preliminary checklist.

The draft ICARS rating system was disseminated to a consensus panel consisting of 16 surgeons (10 expert robotic surgeons and 6 expert laparoscopic surgeons with an intermediate proficiency in robotic surgery). Surgeons were recruited to the panel by invitation on the basis of their experience in surgical training and development. A Delphi process was used to refine the checklist and identify the key relevant behaviours for NTS assessment. In the first round, the comprehensive checklist of behaviours was distributed. Panel members were asked to rate the importance of each component for checklist inclusion using a 5-point Likert scale (5- definitely important to 1- definitely not important). Inclusion was determined by a mean score of $\geq 4$. Excluded components were removed and the Delphi process continued until there was saturation of information.
The approved list was then categorised and formatted into an appropriate assessment template. A five-point rating scale was chosen to rate the specific behaviours matched to subjective standards to aid assessment. These markers ranged from 1 = unacceptable to 5 = excellent. This scale was chosen to ensure raters had specific scope for differentiating subjects. The finalised ICARS rating system was then recirculated amongst the experts. All members of the consensus panel approved the finalised ICARS rating system. Expert opinion on the assessment tool was formally gathered using a qualitative and quantitative questionnaire.

8.2.2 Validation Protocol

An observational trial was undertaken to validate the ICARS behavioural rating system. Participants were invited to take part in the study as part of the trial described in Chapter 9, p. 185. Participants were recruited through open invitation with no specific selection criteria were set. Participants were grouped according to their surgical experience. Standards of surgical experience were based in review of data on learning curves as discussed in previous chapters. Novices were defined as having no surgical operative experience and less than 4 hours of robotic or laparoscopic simulation training. Intermediate proficiency was set at between 1 and 50 robotic cases independently whilst the expert standard was having performed over 50 robotic cases independently.

Within the simulated operating room environment, participants completed a surgical task and their NTS were assessed. Principle analysis involved the assessment of 59 robotic novices. In addition, intermediate and expert robotic surgeons completed the study. Prior to the study all novice participants were given a 1 hour hands-on-training session on robotic suturing and didactic training
on performing a urethrovesical anastomosis. The assessment task comprised a urethrovesical anastomosis using a synthetic dry-lab model (3-Dmed, OH, USA).

An “igloo” distributed operating room was used to authentically recreate the OR environment[119, 31]. Two actors played the roles of scrub assistant and anaesthetist. Auditory and visual cues were used to further increase the realism. OR trolleys, drapes and other equipment were used to portray a realistic operating room environment. Audio recordings of background OR noise were used to create the correct ambient auditory environment. Identical scripted scenarios were used for all participants and each scenario was limited to 15min. Whilst completing the suturing, participants were exposed to four scripted stressor events of increasing magnitude. Firstly, the participants were engaged in direct conversation. Then distracting music at a high volume was played before the participants were questioned on the state of the patient and potential bleeding. Finally, the patient was simulated to become haemodynamically unstable, which the surgeon was required to manage the scenario whilst completing the procedure. Each scenario was scripted to test a variety of NTS. For example the first scenario prompted interaction between the surgeon and theatre team testing social skills such as communication and team working. Scenario two assessed aspects of decision making, situational awareness and leadership. Scenario three assessed all the principle NTS domains including leadership, communication and teamwork, decision making and situational awareness. All scenarios were recorded using 3 external rooms cameras in addition to the internal video feed from the robotic endoscopic camera. For the principle analysis, all videos were blindly evaluated post hoc by a panel of expert robotic surgeons and an NTS expert. All participants were assessed by the NTS expert and nested within the expert robotic surgeons with overlap to enable agreement analysis. Secondary analysis of intermediate and expert robotic participants was undertaken by expert robotic surgeons only. All participants were assessed using ICARS and NOTSS.
8.2.3 Statistical Analysis

Categorical data was reported as frequency, n, and percentage. Continuous data was reported as mean + standard deviation (SD). Given its widespread implementation and validation, NOTSS was defined as the gold standard. Correlation with ICARS was evaluated using a Bland-Altman plot. 95% confidence intervals were corrected for the variance and bias associated with repeated observations[25]. To allow for direct comparison, mean NOTSS and ICARS values were normalised using Z scores. The ability of ICARS to accurately discriminate between participants assessed through the comparison of novice, intermediate and expert participants using the Kruskal Wallis Test. For the purposes of this study we have assumed a direct correlation between surgical experience and technical ability[35]. Global interrater reliability was assessed using Krippendorff’s α given the nested rating design. Direct comparison of raters was performed using an intraclass correlation coefficient for agreement (ICCagreement) coefficients with a two-way random effects model. Internal consistency of the principle domains was tested using Cronbach’s α. 0.7 was broadly defined as the good minimum level for internal consistency[62] however all final decisions were made through consensus with the expert panel. Floor effects were considered present if more than 15% of the novices achieved the minimum score[151]. Ceiling effects could not be assessed given the relatively low numbers of expert participants.

8.3 Results

8.3.1 Development of ICARS

During the first phase of development, 45 distinct behaviours constituting robotic NTS were identified and included in the preliminary checklist. The response rate
for the Delphi process was 100%. Two rounds of the Delphi process were required to reach a consensus amongst the panel and produce the finalised list of NTS behaviours. The final checklist contained 28 core component behaviours, divided into four domains and seven categories (Figure 8.1, p. 171). In addition to generic domains of situation awareness, decision making, task management, leadership and communication and team work, three further domains were felt to warrant inclusion (WHO checklist completion, Console Set Up and Stress and Distractors).
### Interpersonal and Cognitive Assessment for Robotic Surgery (ICARS)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Category</th>
<th>Component</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checklist and Equipment</td>
<td>Checklist</td>
<td>Completes WHO surgical checklist</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td>Console</td>
<td>Appropriate robot settings and console adjustments set</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td>Interpersonal Skills</td>
<td>Communication &amp; Team skills</td>
<td>Effective verbal communication whilst at the console</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate interaction with bedside assistant surgeon</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate interaction with anaesthetist and theatre staff</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fringes/ initiates in non-invasive feedback with theatre staff</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td>Leadership</td>
<td>Instructs the team accordingly and politely</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Effective management of workload and resources</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Co-ordination of activities and team from console</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Co-ordination of activities and team whilst at patient bedside</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comfortable delegating tasks to team members</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maintenance of professional standards</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td>Cognitive Skills</td>
<td>Decision-Making</td>
<td>Appropriate decision in the event of equipment failure</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate decisions made at the bedside</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prompt diagnosis of unforeseen/ unexpected patient event</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fast decision making in emergency situation</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Generation, selection and implementation of solution option(s)</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td>Situational Awareness</td>
<td>Outcome review of management decision</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Awareness of patient status throughout the procedure</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ability to deal with patient at the bedside when necessary</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ability to adapt quickly if a problem arises</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anticipation of potential problems/ difficulties</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Role awareness of surrounding team members whilst at the console</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td>Resource Skills</td>
<td>Stress and Distractors</td>
<td>Understands personal limitations and asks for help if necessary</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identification of stressor/distraction</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maintenance of cognitive and interpersonal skills</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maintenance of technical skills</td>
<td>N/A 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td>Scoring Key</td>
<td>[N/A] Not Applicable</td>
<td>1 Unacceptable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 Poor</td>
<td>3 Acceptable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 Good</td>
<td>5 Excellent</td>
</tr>
<tr>
<td>Overall Score</td>
<td></td>
<td>1 Unacceptable</td>
<td>2 Poor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 Acceptable</td>
<td>4 Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 Excellent</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 8.1 The ICARS Rating System
8.3.2 Validation of ICARS Agreement

Feasibility, Acceptability and Educational Impact

All of the expert panel agreed that the use of ICARS for NTS training was important for training in robotic surgery with 86% agreeing that ICARS could be appropriately applied to the robotic surgical environment (Figure 8.2, p. 173). Interestingly this was in the context of only 53% having ever used an alternative NTS checklist. 68% of the panel did not think that NTS were currently recognised within robotic surgical training. 73% of the experts believed that ICARS would be straightforward to use with only 40% felt an additional guidance sheet would be required prior to use, highlighting its easy-to-use design. There was significant support for the implementation of ICARS both as a learning and assessment tool, 80% agreed that NTS could be actively assessed using ICARS and help identify deficits in NTS. 73% believed ICARS would also promote beneficial discussion and learning follow the case. As a result, 80% of experts agreed that they would be happy to implement ICARS during training.

Validation and Reliability Assessment

73 participants completed the trial consisting of 59 novices, 6 intermediate surgeons and 8 expert robotic surgeons. The novices had an average of 30min simulation experience on open or laparoscopic simulation. None had robotic experience. The intermediate and expert surgeons had mean robotic experience of 5 and 430 cases respectively. In contrast the laparoscopic and robotic simulation experience between intermediates and experts were statistically similar. Variations in NTS training experience were also noted; experts had greater experience but this did not reach significance (Table 8.1, p. 174).

Comparing result of assessment using ICARS and NOTSS, a high degree of correlation was seen on Bland-Altman plot (Figure 8.3, p. 175). The Bland
Fig. 8.2 Expert Assessment of Feasibility, Acceptability and Educational Impact

* Scored on Likert Scale 1 (lowest) - 5 (highest)
### Table 8.1 Demographic Data and Surgical Experience of Validation Study Participants

<table>
<thead>
<tr>
<th></th>
<th>Novice</th>
<th>Intermediate</th>
<th>Expert</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>59</td>
<td>6</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td>22 ± 3</td>
<td>41.2 ± 5.3</td>
<td>45.2 ± 9.3</td>
<td>-</td>
</tr>
<tr>
<td>Gender (%male)</td>
<td>66</td>
<td>66</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Total Years of Medical Training</td>
<td>3.6 ± 2.0</td>
<td>22.2 ± 5.3</td>
<td>26.8 ± 9.6</td>
<td>-</td>
</tr>
<tr>
<td>Robotic Experience (cases)</td>
<td>0</td>
<td>5.8 ± 9.7</td>
<td>436 ± 697</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Laparoscopic Experience (cases)</td>
<td>0</td>
<td>103.3 ± 85.0</td>
<td>234.4 ± 351.2</td>
<td>1</td>
</tr>
<tr>
<td>Robotic Simulation Experience (hours)</td>
<td>0</td>
<td>16.2 ± 20.3</td>
<td>18.1 ± 14.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Non Technical Skills Training Experience (hours)</td>
<td>0</td>
<td>4.0 ± 4.3</td>
<td>10.0 ± 8.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Altman analysis indicates an appropriately narrow 95% confidence interval (z score -0.57 to 0.55) with the majority of plots falling within. The graph demonstrates a uniform scatter of plots suggesting good agreement. Bias was close to zero (-0.007).

![Bland-Altman Analysis Comparing ICARS with NOTSS](image)

Fig. 8.3 Bland-Altman Analysis Comparing ICARS with NOTSS

The ability to accurately differentiate differing levels of NTS competence was assessed by the comparative analysis of novice (n=59), intermediate (n=6) and expert (n=8) surgeons. The assumption that NTS skill positively correlates with surgical experience was verified through our analysis. Significant differences in ICARS score were shown between the 3 groups with experts performing best followed by intermediate participants, p=0.01 (Figure 8.4, p. 176).

Across all raters Krippendorff’s $\alpha$ was 0.42 indicating a moderate agreement[129]. However, when compared directly, a greater degree of agreement between raters was seen with a mean ICC of 0.60. Comparison of ratings from the NTS expert and expert robotic surgeons showed a high degree of agreement (ICCagreement = 0.70).
Fig. 8.4 Comparison of ICARS Between Novice, Intermediate and Expert Participants
Assessment of the internal structure showed the checklist was found to be reliable (Table 8.2, p. 179). All five multi-component categories (Communication; Team Skills; Leadership; Decision Making; Situational Awareness; Stress and Distractors) had high $\alpha$ coefficients (median = 0.92 range 0.85-0.94) demonstrating that the questions accurately represent the category constructs in each case. On analysis of the individual components, only the components related to bedside did not meet the predetermined standard for category construct (Cronbach $\alpha$ 0.70). Of these only one, “Appropriate Interaction with bedside assistant surgeon”, would result in an improved $\alpha$ for the category if deleted. Following discussion amongst the authors, it was decided to retain all these components including “Appropriate Interaction with bedside assistant surgeon”, given the importance of including bedside behaviour assessment in robotic NTS assessment. No floor effects on found on analysis of novice participants results further supports content of ICARS.

Table 8.2 ICARS Inter Item Correlation

<table>
<thead>
<tr>
<th>Domain</th>
<th>Category</th>
<th>Component</th>
<th>Category $\alpha$ Coefficient</th>
<th>Corrected Component-Tot Correlation</th>
<th>$\alpha$ if Component deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpersonal Skills</td>
<td>Communication &amp; Team skills</td>
<td>Effective verbal communication whilst at the console</td>
<td>0.85</td>
<td>0.72</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate Interaction with bedside assistant surgeon</td>
<td></td>
<td>0.56</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate interaction with anaesthetist and theatre staff</td>
<td></td>
<td>0.80</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Engages/ initiates in confirmatory feedback with theatre staff</td>
<td></td>
<td>0.74</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Instructs the team accordingly and politely</td>
<td></td>
<td>0.74</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 8.2 (continued)
Table 8.2 ICARS Inter Item Correlation

<table>
<thead>
<tr>
<th>Domain</th>
<th>Category</th>
<th>Component</th>
<th>Category α Coefficient</th>
<th>Corrected Component-Total Correlation</th>
<th>α if Component deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective management of work-load and resources</td>
<td>0.84</td>
<td></td>
<td></td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Co-ordination of activities and team from console</td>
<td>0.85</td>
<td></td>
<td></td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Co-ordination of activities and team whilst at patient bedside</td>
<td>0.60</td>
<td></td>
<td></td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Comfortable delegating tasks to team members</td>
<td>0.77</td>
<td></td>
<td></td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Maintenance of professional standards</td>
<td>0.81</td>
<td></td>
<td></td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>Appropriate decision in the event of equipment failure</td>
<td>0.81</td>
<td></td>
<td></td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Appropriate decisions made at the bedside</td>
<td>0.66</td>
<td>0.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prompt diagnosis of unforeseen/unexpected patient event</td>
<td>0.89</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fast decision making in emergency situation</td>
<td>0.90</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generation, selection and implementation of solution option(s)</td>
<td>0.77</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome review of management decision</td>
<td>0.85</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness of patient status throughout the procedure</td>
<td>0.78</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to deal with patient at the bedside when necessary</td>
<td>0.66</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability to adapt quickly if a problem arises</td>
<td>0.75</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anticipation of potential problems/difficulties</td>
<td>0.82</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role awareness of surrounding team members whilst at the console</td>
<td>0.76</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8.2 (continued)
8.4 Discussion

This study demonstrates the successful development of the first NTS behavioural marker system specifically for robotic surgery. Extensive evidence has been reported for the construct validity of ICARS, supporting the use of ICARS scores to assess NTS in robotic surgery.

ICARS content was derived both through observational data from live robotics surgery as well as expert opinion through a Delphi consensus approach supporting the content validity of the tool. This approach identified 28 key behaviours relevant to robotic surgery which organised into a structured behavioural rating system. Response process analysis is particularly important for NTS assessment where both target behaviours and assessment tools used to measure there are relatively subjective when contrasted to specific outcomes in technical skills assessment. This study’s response validity was supported by the use of video

Table 8.2 ICARS Inter Item Correlation

<table>
<thead>
<tr>
<th>Domain</th>
<th>Category</th>
<th>Component</th>
<th>Category α Coefficient</th>
<th>Corrected Component-Total Correlation</th>
<th>α if Component deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource Skills</td>
<td>Stress, Fatigue and Distractors</td>
<td>Understands personal limitations and asks for help if necessary</td>
<td>0.94</td>
<td>0.72</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identification of stressor/distraction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maintenance of cognitive and interpersonal skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maintenance of technical skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Professional and appropriate choice of resolution</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The End
recordings of all participants to enable unbiased, accurate performance assessment. Furthermore whilst rater assessment was not formally undertaken, qualitative feedback demonstrated that raters felt that the ICARS tool was easy to use irrespective of whether the raters had undergone prior NTS assessment training. Internal structure of ICARS was extensively validated. All five multi-component categories were found to have very strong relationship with their constituent components. Furthermore, there was high internal consistency of all component items aside from those relating to bedside behaviours. All individual components bar one (“Appropriate Interaction with bedside assistant surgeon”) were also shown to have a positive effect on overall consistency as demonstrated by remove-one analysis. Rater analysis also demonstrated a high degree of consistency between raters despite a large pool of raters. Only eight (53%) had previous experience and training in NTS assessment (NOTSS in all cases) but the majority of raters found the rating system easy to use. Consequently the study does support the internal structural validity of ICARS with scores shown to be generalisable, reliable and reproducible. Given its widespread validation and application to surgical training, NOTSS was selected as the gold standard comparator[264, 171]. A high degree of agreement was seen between ICARS and NOTSS with low bias. Alongside this, a positive correlation was also seen between experience (defined as expert, intermediate and novice surgeons) and ICARS score. This is predicated on previous research showing a correlation between surgical experience and the level of NTS[35].

The use of a simulated scenario to validate ICARS offers a number of benefits. Distributed simulation has been shown to be effective in NTS assessment and enabled standardised, fully immersive training conditions to be set for all 73 participants[221, 31]. Post-hoc video analysis allowed blinded assessment to be undertaken by the panel of robotic and NTS experts. Combined with the high number of participants, we were able to conduct a comprehensive validation
assessment of the ICARS behavioural rating tool. Interestingly this study also highlighted the lack of NTS training currently seen in robotic surgery with almost half the surgeons never having used an NTS checklist. A key development aim of this project was to create a tool that is easily used in the everyday setting. An overall structure similar to establish task specific rating systems was used to maximise familiarity for surgeons[269, 74, 155]. In addition, more specific detail was given to guide accurate behavioural assessment. Both the feedback from the expert panel and the high degree of agreement found between the experts with and without specialist non-technical skills experience, supporting its usability.

Robotic surgery poses considerable NTS challenges to the surgeon distinct to those of open or even laparoscopic surgery. Face-to-face interaction between the surgeon and his team is greatly reduced and there is an increased reliance on the surgical assistant and scrub team. Furthermore, the surgeon’s role changes during the operation as he moves between the bedside and console. Each requires differing NTS. These differences were reflected both in the NTS selected during the development of ICARS and in the validation results. ICARS assesses the surgeon both at the bedside and the console. Additionally, ICARS includes a number of unique behaviours not described in generic rating systems including awareness of team members whilst at the console, awareness of the patient status and equipment failure[269, 245, 154, 206]. These NTS idiosyncrasies of robotic surgery result from the both the surgeon’s reliance on the robotic system and his detachment from the operating room when sat in the encompassing environment of the console. The unique nature of these skills is further highlighted by comparison of novice, intermediate and expert performances. Experts were substantially better at coordinating the activities from the console, appropriate communication, communicating with the team and managing equipment failure (Appendix E). Interestingly intermediate participants were better at managing the patient at the bedside which may be reflected by their greater experience as bedside assistants. It
is important to acknowledge that the only lower scoring components of the rating system related to bedside assistance. This may reflect the assessment tasks which focussed primarily on the console surgeon. Further work needs to be undertaken to specifically evaluate the role of ICARS in assessing bedside performance.

Validation does need to be considered in light of the studies limitations. Use of a simulated OR was appropriate for initial validation however ongoing evaluation of ICARS during live surgery and in specialities other than urology is still required. Similarly, although we have demonstrated the reliability of ICARS om comparison to NOTSS, the gold standard, continuing evaluation is required to measure test stability over time as well as to determine appropriate benchmarks for training[7]. Assessment using ICARS will provide an objective measure of NTS that can be used alongside technical skills assessments already include in robotic surgical training[72]. In this study, participants were categorised into three cohorts dependent on surgical experience. Further assessment needs to be undertaken to investigate whether participants within these categories can also be reliably differentiated. Like other behavioural rating systems, ICARS will also be important component of training for structured debrief.

ICARS has been developed as a rating tool to provide objective and structured assessment of NTS during robotic surgery. As the use of robotic surgical systems continue to grow, integration of ICARS within training programmes will provide objective and evidenced-based evaluation of NTS in robotic surgery. The use of ICARS will support structured NTS training and the standardised assessment it provides will enable further research into improving safety and performance in robotic surgery. Currently the only commercially available surgical robot is the Da Vinci System, which this study was based on. The key NTS behaviours identified during this study that comprise ICARS will be equally applicable to the new robotic systems currently in development. As a result, ICARS will be
applicable to all robot training, allowing surgeons to directly compare and assess their NTS irrespective of the specific robotic system used.

8.4.1 Summary

This chapter reports the development of the first NTS behavioural rating system for robotic surgery. A structured, evidenced-based approach was applied building on previous studies developing NTS rating tools for other specialities. A comprehensive taxonomy of NTS behaviours relevant to robotic was compiled using interviews with expert surgeons and observation of live surgery. Using a Delphi methodology, a panel of 16 expert surgeons were consulted to identify behaviours important to NTS assessment. These behaviours were organised into an appropriate assessment template. Experts were consulted on the feasibility, applicability and educational impact of ICARS. An observational trial was used to validate ICARS. 73 novice, intermediate and expert robotic surgeons completed a urethrovesical anastomosis within a simulated operating room. NTS were tested using four scripted scenarios of increasing difficulty. Performances were video recorded. Robotic and NTS experts assessed the videos post-hoc using ICARS and NOTSS. 28 key non-technical behaviours were identified by the expert panel. The finalised behavioural rating system was organised into 4 principle domains and 7 categories. Expert opinion strongly supported its implementation. ICARS was found to be equivalent to NOTSS on Bland-Altman analysis and accurately differentiated between novice, intermediate and expert participants, $p=0.01$. Moderate agreement was found between raters, Krippendorff’s $\alpha = 0.4$. The internal structure of ICARS was shown to be consistent and reliable (median Cronbach $\alpha = 0.92$, range 0.85-0.94). Initial validation of the ICARS rating systems demonstrates it to be an effective and reliable tool for assessing NTS during robotic surgery. Ongoing evaluation of ICARS during simulation and
real-life training will help guide the development of a structured training protocol for NTS in robotic surgery.
A Randomised Controlled Trial of Motor Imagery Training for Technical and Non-Technical Skills in Robotic Surgery

9.1 Introduction

The advantages of proficiency based training in driving skill acquisition over historical models of experiential learning are widely recognised in medicine[94]. Whilst experience is often still considered synonymous with expertise, increasingly goal-directed, focussed training forms the basis of curricula across the spectrum of specialities[95, 69]. Expansion of simulation training has allowed adoption of the deliberate practice model of training centred on repeated, effortful practice directed through feedback to improve performance[68]. The wide variety of simulation tools now available has enabled training to move towards a more individualised, proficiency-based approach. However physical simulation is not the only approach available for surgical training. Mental training in the form of motor imagery is used as an important training adjunct in a variety of fields ranging from sport and music to rehabilitation medicine[216]. Motor imagery (MI) is defined as the mental rehearsal of a motor task without physical execution of the movements
involved[159]. MI is based upon a theory of functional equivalence with motor imagery and motor preparation and execution sharing a common neural substrate. Motor imagery results in similar neural activity to motor planning and activity and sustained MI practice results in improved motor performance of the task. Similar physiological responses such as heart rate and electromyography have also been found to occur in response to both MI and motor execution[169, 227].

MI is affected by a variety of variables and factors that result in different forms of neural activation. For example visual MI whereby one simply imagines a motor task being performed either in the first person or third person results in different patterns of activation compared to kinaesthetic imagery whereby one imagines the whole sensation of performing the motor task[153]. In this example, for motor learning, kinaesthetic MI has been found to be more effective[159]. Studies have identified a variety of factors that influence both MI and its effects on motor performance[153, 88].

Yet in contrast to elite sports and other performance-based roles, the potential for such cognitive training to confer further educational and performance gains in surgery remain largely unexploited[255, 220].

Alongside offering numerous advantages, minimally invasive surgery (MIS) poses new technical and non-technical skill (NTS) challenges to the surgeon. Operative differences such as the reduction or loss of haptic feedback cause difficulties particularly for trainee surgeons[233, 42]. The minimally invasive surgeon is also confronted by new NTS challenges due to the alterations to the working environments and team setup. Specific NTS training is vital to ensure the surgeons continue to operate safely and effectively[82]. Physical simulators have been extensively incorporated into training curricula, particularly for minimally invasive surgery, to meet these new training demands. The objective of this study is to determine whether evidence-based MI can successfully be used for training technical and non-technical skills in minimally invasive surgery.
9.2 Methods

A single blinded, parallel group randomised controlled trial was conducted. Participants were recruited from three large London hospital trusts and four associated medical schools by open invitation between May 2015 and August 2015. Recruitment was open to junior doctors and medical students who had completed at least one year of clinical training. Novice level was defined as having no prior experience in performing MIS either within an operating room or simulation. More senior trainees (trainee surgeons and above) were excluded from the trial to ensure an homogenous baseline ability across both groups.

9.2.1 Trial Protocol

The trial was conducted in the Vattikuti Institute of Robotic Surgery, King’s College London. Participants were required to attend three separate sessions (Figure 9.1, p. 188). Initially all participants completed a basic skills training course. This consisted of didactic and hands-on-training in basic robotic skills and robotic suturing. Robotic training was performed using a virtual reality simulator, the Da Vinci Surgical Simulator (DVSS) attached to the Da Vinci Xi Surgical Robot (Intuitive Surgical Inc, Sunnyvale, CA, USA). All participants completed two basic robotic skills modules (Pick and Place; Pegboard 2) in addition to a robotic suturing module (Interrupted Suturing). Inbuilt proficiency scores from the simulator were used to guide training. The score from the initial attempt of Pegboard 2 provided an objective baseline measure of competency.
Fig. 9.1 MI RCT Study Protocol
Following successful completion of the basic skills training course, simple randomisation was performed using a random number generator to assign participants to either cognitive training or control (GraphPad Software, Inc. La Jolla, CA, USA). Randomisation and allocation of participants was concealed from the researcher enrolling participants to the further trial sessions. Given the nature of the trial, blinding of the participants or trial organisers was not possible. The independent expert surgeons evaluating the final skills assessment were blinded to the intervention assignment. Participants were invited to return for further training in groups of four according to their randomisation status. The assessment task selected was urethrovesical anastomosis (UVA). This specific procedure was selected to provide the necessary degree of complexity to differentiate participants whilst also being accurately simulated using a synthetic, dry lab model (3-Dmed, OH, USA). Prior to training all participants in both arms were shown a 30min video explaining the steps for UVA using the same synthetic model. The control group underwent standard training which comprised a further instructional video on the correct technique for the UVA. The intervention group underwent MI training. Training involved mentally rehearsing a UVA with the aid of a MI script. Initially participants being taken through the MI script with an experienced trainer. Following this training session participants then continued their MI training in their own time. The MI script was developed in collaboration with experts in MI and expert robotic surgeons using the PETTLEP model[103]. The model provides a structured checklist to address key imagery processes for successful imagery. PETTLEP is an acronym identifying the seven domains for developing an effective motor imagery script (Physical, Environment, Task, Timing, Learning, Emotion, Perspective). The MI checklist was developed according to this protocol. The checklist instructed participants to complete the MI training in the same seated position as at the robotic console (Physical). Training was initially undertaken in the same setting (Environment). Participants were asked to imagine completing
all the steps of the UVA task, focussing on specific technical factors for example ensuring correct rotation of the wrist to pass the needle cleanly through the tissue (Task). Participants were instructed to imagine the task in real time (Timing) and from a first-person Perspective (Imagining the action through one’s own eyes). The checklist also provided a number of sensory triggers to encourage the accurate Emotional imagery of the task. These domains were emphasised to the participants to stimulate vivid and effective imagery.

During the third session, participants underwent identical technical and non-technical skill assessment. First participants completed a warm-up exercise using the Da Vinci Xi surgical robot to complete a basic skills dry-lab training model (Pea on a Peg, 3-Dmed, OH, USA) before completing a practice attempt of an UVA (not assessed). Following this warm up, participants underwent assessment within a simulated operating room (OR). An inflatable, simulated OR, the Igloo, was used to accurately recreate the OR environment with actors playing the roles of scrub nurse and anaesthetist. A variety of auditory and visual cues were used to further increase the realism. The actors followed identical scripted scenarios for all participants. Each scenario was limited to 15min. Whilst completing the UVA task, participants were exposed to three stressor events to test their NTS. These increased in difficulty, ranging from being engaged in simple conversation with the surgical team to the patient becoming haemodynamically unstable. Participant performances were recorded using both the video feed from the Da Vinci endoscopic camera as well as three external room video cameras. The video recorded scenarios were blindly assessed post-hoc by four expert robotic surgeons (defined as having performed over 200 robot assisted radical prostatectomies independently) and a surgical non-technical skills expert. A nested design was used whereby all participant videos were assessed by two raters. The primary outcome measures of the study were technical and non-technical skill. Technical skill was assessed using Global Evaluative Assessment of Robotic Skills (GEARS)
whilst the Non-Technical Skills for Surgeons (NOTSS) behavioural rating system was used to rate NTS. The secondary outcome measure was the quality of MI. This was evaluated using the Movement Imagery Questionnaire (MIQ), a validated subjective assessment of MI quality, modified for robotic surgery[56, 110, 15].

9.2.2 Power Calculation

Power calculations were based on a prior studies of MI training using global rating scale (GRS) to assess performance[110, 127, 258, 141]. To detect significance at $p = 0.05$, a 15% difference in the mean GRS following training with a power of 80% assuming a 10% drop out rate, required a minimum sample size of 53 was required.

9.2.3 Statistical Analysis

Data was found to have a Gaussian distribution. Between group differences were analysed using independent samples T-tests. Inter-rater agreement was measured using Krippendorff’s alpha. Strength of association between baseline variables and mean GEARS scores was measured using the Pearson correlation coefficient. Analysis of covariance (ANCOVA) was used to control for the effects of potential covariate effects. Statistical analyses were conducted using SPSS v23 (Chicago, Illinois). Graphs were created using Prism version 7.0 (GraphPad Software, Inc. La Jolla, California, USA).

9.3 Results

90 participants responded to the initial recruitment invitation and 77 completed the basic training programme. Of these 64 returned for randomisation and completed the full study. Videos of two participants in the control arm were
corrupted and could not be assessed. Per protocol analysis comprised data from 33 participants randomised to MI training and 29 to standard training (Figure 9.2, p. 193). The two groups were well matched in terms of basic demographics, and clinical experience.
Fig. 9.2 MI RCT Consort Flow Diagram
Table 9.1 Participant Demographic Details, Clinical Experience and Baseline Skill

<table>
<thead>
<tr>
<th></th>
<th>MI Group</th>
<th>Control Group</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Gender, n (%)</td>
<td>20 (61%)</td>
<td>17 (59%)</td>
<td>0.57</td>
</tr>
<tr>
<td>Age, mean years ± SD</td>
<td>22.3 ± 2.5</td>
<td>23.8 ± 3.9</td>
<td>0.07</td>
</tr>
<tr>
<td>Clinical Experience, mean years ± SD</td>
<td>3.6 ± 2.0</td>
<td>3.9 ± 2.1</td>
<td>0.54</td>
</tr>
<tr>
<td>MIS Surgical Experience, mean cases assisted ± SD</td>
<td>2.6 ± 9.3</td>
<td>1.3 ± 3.1</td>
<td>0.38</td>
</tr>
<tr>
<td>Baseline Technical Ability, combined overall score ± SD</td>
<td>1238 ± 429</td>
<td>1356 ± 497</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Following initial VR training, baseline MIS ability measured using the DVSS was found to be equivalent between groups (Table 9.1, p. 194). Interrater reliability between raters of technical and non-technical skills assessment was high, Krippendorff’s $\alpha = 0.85$.

### 9.3.1 Technical Skill Assessment

Comparison of technical performance using the total GEARS showed MI training to be significantly superior to standard training alone (Figure 9.3, p. 195). MI training resulted in a mean ± SD GEARS score of 13.1 ± 3.25 compared to 11.4 ± 2.97 following standard training (p = 0.03).
Fig. 9.3 Technical Performance of MI Training and Control Groups
Table 9.2 Comparative Analysis of GEARS Subscale Score for MI and Control Groups

<table>
<thead>
<tr>
<th>GEARS Subscale</th>
<th>Intervention</th>
<th>Mean Score ± sd</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth Perception</td>
<td>Control</td>
<td>1.93 ± 0.55</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.19 ± 0.67</td>
<td></td>
</tr>
<tr>
<td>Bimanual Dexterity</td>
<td>Control</td>
<td>2.00 ± 0.68</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.48 ± 0.66</td>
<td></td>
</tr>
<tr>
<td>Efficiency Control</td>
<td>Control</td>
<td>1.75 ± 0.53</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.07 ± 0.65</td>
<td></td>
</tr>
<tr>
<td>Force Sensitivity</td>
<td>Control</td>
<td>1.90 ± 0.59</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.07 ± 0.64</td>
<td></td>
</tr>
<tr>
<td>Autonomy</td>
<td>Control</td>
<td>1.90 ± 0.54</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.07 ± 0.80</td>
<td></td>
</tr>
<tr>
<td>Robotic Control</td>
<td>Control</td>
<td>1.92 ± 0.68</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.23 ± 0.52</td>
<td></td>
</tr>
</tbody>
</table>

Detailed analysis of the constituent subscales is shown in Table 9.2, p. 196. Significant improvements in performance were seen in 3 of the 6 domains assessed by GEARS; Bimanual Dexterity, Efficiency and Robotic Control. Analysis of potential covariates showed weak correlations between mean GEARS and both baseline technical ability (r = 0.27, p = 0.051) and clinical experience (r = 0.24, p = 0.06). MI training effects remained the most significant factor in determining technical performance after controlling for the effects of these potential confounders (p = 0.016, F (1,50) = 6.21).

### 9.3.2 Non-Technical Skills Assessment

NTS performance did not differ significantly between MI and standard training participants (mean total NOTSS score ± SD 25.8 ± 7.34 vs 26.4 ± 9.13 respectively, p = 0.77). No demonstrable differences were found either in overall NOTSS scores or individual subcategory scores between the two groups (Table 9.3, p. 197).
9.3 Results

Table 9.3 Comparative analysis of NOTSS subscale score for the MI and Control Groups

<table>
<thead>
<tr>
<th>NOTSS Subscale</th>
<th>Intervention</th>
<th>Mean Score ± sd</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situational Awareness</td>
<td>Control</td>
<td>2.21 ± 0.76</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.15 ± 0.62</td>
<td></td>
</tr>
<tr>
<td>Decision Making</td>
<td>Control</td>
<td>2.26 ± 0.78</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.26 ± 0.62</td>
<td></td>
</tr>
<tr>
<td>Communication and Teamwork</td>
<td>Control</td>
<td>2.12 ± 0.74</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.02 ± 0.63</td>
<td></td>
</tr>
<tr>
<td>Leadership</td>
<td>Control</td>
<td>2.19 ± 0.78</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>2.18 ± 0.65</td>
<td></td>
</tr>
<tr>
<td>Mean Total NOTSS Score</td>
<td>Control</td>
<td>26.4 ± 9.13</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>25.8 ± 7.34</td>
<td></td>
</tr>
</tbody>
</table>

9.3.3 Motor Imagery Assessment

54 of the participants completed the post assessment MIQ. Overall MI training resulted in a significantly higher movement imagery score than standard training (p = 0.04). Greater motor imaging ability was demonstrated across the majority of the MIQ nodal points but these did not reach significance (Table 9.4, p. 198).
<table>
<thead>
<tr>
<th>MIQ Questions</th>
<th>Intervention</th>
<th>Mean Score ± sd</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) How ready do you feel to carry out a UVA</td>
<td>Control</td>
<td>4.08 ± 1.81</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>4.72 ± 1.47</td>
<td></td>
</tr>
<tr>
<td>2) How confident do you feel to carry out a UVA</td>
<td>Control</td>
<td>3.99 ± 1.72</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>4.66 ± 1.43</td>
<td></td>
</tr>
<tr>
<td>3) How well do you think you can perform the procedure compared to others at your stage</td>
<td>Control</td>
<td>4.19 ± 1.69</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>4.85 ± 1.26</td>
<td></td>
</tr>
<tr>
<td>4) How helpful is the MI activity you have been performing in preparing you to perform the robotic procedure</td>
<td>Control</td>
<td>4.84 ± 1.48</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>5.54 ± 1.28</td>
<td></td>
</tr>
<tr>
<td>5) How easily can you see yourself performing a robotic UVA</td>
<td>Control</td>
<td>4.66 ± 1.71</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>5.10 ± 1.21</td>
<td></td>
</tr>
<tr>
<td>6) How clear and vivid are the images in your mind</td>
<td>Control</td>
<td>5.03 ± 1.20</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>5.61 ± 1.27</td>
<td></td>
</tr>
<tr>
<td>7) How easily can you feel yourself performing the procedure</td>
<td>Control</td>
<td>4.51 ± 1.58</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>4.91 ± 1.23</td>
<td></td>
</tr>
<tr>
<td>Overall Mean Score</td>
<td>Control</td>
<td>4.46 ± 1.23</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>MI</td>
<td>5.11 ± 0.98</td>
<td></td>
</tr>
</tbody>
</table>
9.4 Discussion

This single-blinded randomised controlled trial has demonstrated the effectiveness of MI in supporting MIS simulation training. Structured MI training using the PETTLEP model led to a significantly greater improvement in technical performance over standard training with 1.7 mean difference in GEARS score. In contrast, there was no benefit of MI training on NTS performance. Furthermore, MI training was confirmed as being effective in improving the imaging ability of participants.

GEARS is the most widely validated trial for global skills assessment in robot surgery during live surgery and in the simulation setting[85, 260, 105]. It provides a reliable global assessment of robotic technical skill rather and was therefore chosen over procedure specific tools such as the Robotic Anastomosis Competency Evaluation (RACE) which focus on the specific skills required for anastomosis[203]. Improvements in GEARS scores were predominantly found in domains representing tasks associated more with the cognitive challenges of robotic surgery than purely technical skill. With MI training, participants scored higher for depth perception, bimanual dexterity and efficiency. Variations in the effectiveness of MI training according to the type of task have been seen previously. Cognitive tasks exhibit greater training effects than motor or strength exercises[64, 110]. In contrast MI training did not translate into NTS improvements, agreeing with the findings of Louridas et al. [141]. Given the MI training provided in this study focussed on the technical performance of a UVA anastomosis, this negative finding may be as a result of a lack of far transfer of learning in MI[166]. However previously specific MI training in NTS has also not be found to be beneficial which raises the question whether the more abstract nature of NTS prevents effective mental visualisation and training[99].
This trial was designed to provide the best available evidence to evaluate the utility of standardised MI training for complex MIS surgical procedures. Whilst it was performed in a simulated operating room environment, distributed simulation such as the Igloo used in this study has been shown to effectively recreate the operating room environment[31]. It offers high fidelity, immersive simulation enabling the effective assessment of the participants technical and non-technical skills. Likewise, the single point of assessment limits the conclusions that can be drawn as to the degree of benefit derived from MI training. In spite of this, MI imagery training resulted in a global improvement in technical performance. Further investigation is necessary to investigate the effect of MI on the complete learning curve and in other training environments.

Different forms of imagery elicit different patterns of neural activation and result in different training effects[159]. For effective motor skill learning, kinaesthetic imagery in which the feeling of performing the task is mental rehearsed in contrast to visual imagery whereby only a visual representation of the movement is produced[153]. As a result, the content of the MI script is critically important in establishing the correct form of imagery. MIQ results from this study show that the PETTLEP model can be used for effective and reproducible MI training in surgery. A key question in MI training that remains unanswered is the influence of participant experience on MI outcomes. No comparative studies in surgical training have been undertaken but trials with experienced participants have greater rates of success compared to those involving novices[58]. It has been argued that experts are better able to undertake kinaesthetic imagery due to more established motor representations of the action and consequently derive more benefit from MI[170, 257]. In contrast, there is evidence to do suggest that novices demonstrate a greater capacity for improvement through MI training particularly in cognitive tasks[64].
A key focus of this study was the provision of structured motor imagery training. For the first time, a MI script written in accordance with evidence base principles of PETTLEP was effectively applied to surgical training. MIQ results support the use of this model of MI training for inexperienced trainees. Focussing on promoting first person, kinaesthetic imagery rather than 3rd person or visual based imagery enables more effective motor skill learning[89, 39]. Functional MR imaging has shown that MI not only leads to the extended neural activations involved in the complex movements required for MIS which require considerable cognitive and motor control but that MI training results in more efficient and focused neural activation[89, 22].

In medicine there is a growing recognition that the attainment of expert proficiency requires persistent active learning. Simulation training is increasingly being used to enable the repeated and focussed training necessary to gaining technical and non-technical skill. The results of this study provide evidence for the integration of MI training alongside existing simulation training programmes for MIS. In comparison to standard training alone, the addition of MI training was beneficial to the process of technical skill training. Structured MI training can be successfully used in inexperienced surgeons and offers flexible training than can easily be applied to other areas of medicine. Teaching surgeons in how to effectively utilise MI within their current training programmes will promote further gains in attainment and efficiency.

9.4.1 Summary

This study evaluated for the first time the role of motor imagery training for technical and non technical skills training in robotic surgery. 64 novice participants were randomised to MI training or control. Structured MI training was provided using the validated PETTLEP model. Technical and non technical skills were
assessed in a simulated operating room environment. A dry lab UVA task was used to assess technical skills. Participants also had to manage 3 NTS scenarios of increasingly difficulty. Performances were blindly assessed by experts using GEARS and NOTSS, validated scoring systems. The results from this RCT provide evidence for the utility of structured MI to augment simulation training for novice surgeons in minimally invasive surgery. Use of the PETTLEP-based MI script led to better imagery of the procedure and consequently improved technical performance.
Simulation training is now widely accepted as an integral component of surgical training. Yet in practice the degree to which it is applied and incorporated into surgical curricula remains highly variable. Robotic surgery offers a unique opportunity to position evidence-based training method at the centre of the training curricula for this emerging surgical technique.

The aims of my thesis were to evaluate the principles methods of simulation training applicable to robotic surgery to support the development of an evidence-based robotic training curriculum. This research has involved a systematic review of current practices in surgical simulation to identify the key factors that promote or hinder surgical simulation training. Original research was then undertaken to evaluate current modalities of simulation training for robotic surgery, addressing the deficits in current training models and to provide the necessary data to support the development of an evidence based curriculum.
10.1 An Overview of the MARS Project

10.1.1 Contribution of the MARS Project to the Science of Simulation Training

Through the work I have reported in this thesis, I have been able to make a number of contributions to the field of simulation training. To date the work presented in these chapters has been presented to a number of international academic conferences as detailed in Table 10.1, p. 206. Furthermore work from this thesis has been published in peer review scientific journals (see Table 10.2, p. 207). The principle academic achievements from the MARS project have been the development of the first NTS behavioural rating tool for robotic surgery (ICARS), the first evaluation of procedural virtual reality training for robotic surgery and the first RCT demonstrating the role of cognitive training in robotic surgery.

In addition to these academic achievements, the MARS project has been able to contribute directly to establishing simulation training for robotic surgery. Development of benchmark scores for the Mimic DV Trainer supported the creation of a basic robot skills training curriculum. Subsequently it has been incorporated into the EAU Robotic VR hands-on-training courses. All participants aiming to complete further robotic training are required to pass the HOT course before they are eligible for further robotic training held at regional training centres endorsed by the ERUS Education Working Group.

The MARS project has also given me the opportunity to develop the first robotic cadaveric simulation training programme. In cooperation with The Freeman Hospital Newcastle we ran two human cadaveric training courses which I developed in accordance with the principles of simulation training. Following the success of these initial courses, we secured funding from The Urology Foundation to support further training for 5 years. Alongside the cadaveric training courses, we
were also able to develop a programme of mini-fellowships for course participants at one of five regional robotic centre. This advanced robotic training programme continues to run.

10.1.2 Systematic Review of Simulation Based Training Programmes

This systematic review was undertaken to provide a report of the current practice in simulation-based training and identify the positive and negative characteristics of current simulation programmes. The unique approach of this systematic review was to focus on the method of technical simulation training provided. To provide a comprehensive overview of how simulation training is currently being delivered, only studies that provided explicit details on the training model were included. A systematic review was undertaken in line with PRISMA guidelines. The results showed an exponential rise in the number of simulation studies being reported with simulation training being delivered for all principle surgical modalities. Most studies did meet a number of the requirements for effective training. In contrast to concerns over the excess use of medical students to test simulation training methods and models, as highlighted by Cook and Hatala [52], medical students and junior doctors formed a minority of course participants. This was particularly applicable to the more advanced, procedural based training courses which were populated with surgical trainees and consultant surgeons. Most training included in this review was delivered as mass training events at fixed times and set durations. This is opposed to the ideal model of distributed training, an ongoing process tailored around an individual’s training needs and other commitments[23]. This may be understood from a logistical point of view especially when new methods of training are being assessed. Widespread use of proficiency-based targets in the reported studies does offer the potential for an expansion of distributed, trainee
### Table 10.1 Results from the MARS Project Presented to Academic Meetings

<table>
<thead>
<tr>
<th>Presentation Title</th>
<th>Academic Meeting</th>
<th>Associated Thesis Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Institutional Validation and Assessment of Training Modalities in Robotic Surgery: The MARS Project[188, 189]</td>
<td>EAU Congress 2017; AUA Congress 2016</td>
<td>n/a</td>
</tr>
<tr>
<td>Developing benchmark scores for the eau hands-On Training (HOT) course in robotic surgery[190]</td>
<td>AUA Congress 2016</td>
<td>3</td>
</tr>
<tr>
<td>Competency based training for virtual reality simulation in robotic surgery[185, 196]</td>
<td>ERUS Congress 2016; WCE 2016</td>
<td>3</td>
</tr>
<tr>
<td>Development of a technical checklist for the assessment of suturing in robotic simulation[90]</td>
<td>AUA Congress 2018</td>
<td>4</td>
</tr>
<tr>
<td>Development and Validation of a Non-Technical Skills Assessment Tool for Robotic Surgery [197, 263]</td>
<td>AUA Annual Congress 2017; ERUS 2016</td>
<td>8</td>
</tr>
<tr>
<td>Training tools for non-technical skills in robotic surgery[185]</td>
<td>ERUS Congress 2016</td>
<td>8</td>
</tr>
<tr>
<td>Content validation of a competency based curriculum for virtual reality simulation in robotic surgery[185, 192]</td>
<td>ERUS Congress 2016; WCE 2016</td>
<td>8</td>
</tr>
<tr>
<td>A randomised controlled trial into cognitive training for technical skills in robotic surgery[185, 186]</td>
<td>ERUS Congress 2016, AUA 2017</td>
<td>9</td>
</tr>
</tbody>
</table>
### Table 10.2 Results from the MARS Project Published in Academic Journals

<table>
<thead>
<tr>
<th>Presentation Title</th>
<th>Journal</th>
<th>Relevant Thesis Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Role of Simulation in Surgical Training[191]</td>
<td>European Urology Focus</td>
<td>1</td>
</tr>
<tr>
<td>Competency based training in robotic surgery: benchmark scores for virtual reality robotic simulation[193]</td>
<td>BJUI</td>
<td>3</td>
</tr>
<tr>
<td>Virtually Competent: A Comparative Analysis of Virtual Reality and Dry-Lab Robotic Simulation Training (accepted for publication)</td>
<td>Journal of Endourology</td>
<td>5</td>
</tr>
<tr>
<td>Development and validation of a tool for non-technical skills evaluation in robotic surgery-the ICARS system[198]</td>
<td>Surgical Endoscopy</td>
<td>8</td>
</tr>
<tr>
<td>Cognitive training for technical and non-technical skills in robotic surgery: a randomised controlled trial.[187]</td>
<td>BJUI</td>
<td>9</td>
</tr>
</tbody>
</table>
centred simulation in the future. Of course the majority also continue to use dry lab simulation tools rather than VR simulators which contrasts with the large body of literature on the effectiveness and reliability of VR simulation. The only exception to this was basic robotic training. For advanced procedural training, wet lab models such as cadavers and live animal simulation were primarily used across most of the surgical modalities. This again contrasts with the limited published validation studies of wet lab simulation models[81]. Objective assessment tools were widely used in the studies but the use of GRS without adequate rater training was highlighted by the relatively poor inter rater reliability. In comparison, less subjective procedural checklists demonstrated higher inter-rater reliability.

This study also revealed some of the limitations of the current literature. Only a minority of trials were high quality according to their MERSQI scores with the majority reporting small participants cohorts from single institutions. Training outcomes, whilst reported in almost all studies, focussed on skills and qualitative feedback from participants with few showing effects on behaviour or patient outcomes. Such deficits have been extensively reported and reflect the difficulty and complexity of studies required to demonstrate empirical for these downstream effects of training.

The results of this study highlight a number of factors important to successful simulation training including inclusion of appropriate participants, use of proficiency based training, objective and validated outcome measures and the need to report on larger multi-centre trials of simulation programmes.

10.1.3 Virtual Reality Simulation Training

The effectiveness of VR training across various surgical disciplines including robotic surgery has previously been demonstrated[3, 19]. The opportunity VR offers to practice specific technical skills has led to it being commonly included in curricula
and training programmes. Results from the systematic review, Chapter 2 (p. 25) have shown, in contrast to other surgical modalities, that VR is used extensively in robotic simulation programmes. For all training, both the simulation tool and the training technique are equally important in ensuring successful training is delivered. In contrast to the relatively extensive validation of VR training tools in robotic surgery[156], training methods to promote effective learning have not been previously evaluated. This study sought to establish the feasibility of competency based training using objective standards in robotic VR training. Partitioning training into sequential tasks of increasing difficulty aims to enhance learning by mirroring the process of motor skill acquisition proposed by Fitts and Posner [73]. Subdividing the learning process in this way ensures participants transition through the three stages, cognitive, associative and autonomous, ensuring that relevant motor skills are developed component by component. Applying these principles to robotic VR training, trainees need to initially gain familiarity with the robotic controls and basic skills, such as clutch control, camera control and endowrist manipulation. Following this, trainees should then apply these skills to more advanced, specific tasks such as knot tying or suturing. This involves both refining their basic skills through practice and error recognition but also gaining new knowledge of the specific tasks. In order for this process to be effective, participants must gain the necessary basic skills before advancing onto more difficult tasks. Assessment plays a key role in ensuring that this process occurs and that teaching has been effective. Another key aspect of simulation training is feedback to promote improvement. Central to the theory of deliberate practice, propounded by Ericsson, is immediate feedback. Comparison of a trainees performance to the expert standard stimulates both reflection and correction of errors[69]. Furthermore, specific goals help to motivate trainees especially in generic tasks that need to appeal to participants with a diverse range of base skills and experience. The results have demonstrated that it is effective and feasible to
use benchmark scores derived from expert performance data to set meaningful targets around which training can be structured.

In this study, I used an evidence based approach based on the training results of 223 participants completing 1565 exercises to both select key simulator exercises and set appropriate benchmark scores. The resultant training programmes comprised 5 exercises; three basic skill exercises and two advanced skill exercises. This programme can be used for both assessment and training. I was able to show that my choice of 75% of the mean expert score provided the right standard against which surgeons can rate their basic robotic skills. Subsequent studies have been able to utilise a similar methodology for establishing scores using other VR simulators[256]. Further work needs to be undertaken to develop benchmark scores for dry lab training to enable training using synthetic models alongside or in place of VR simulation depending on local training infrastructure.

10.1.4 Procedural vs Basic Simulation Training

The initial element of this project was to evaluate the RobotiX Mentor procedural VR simulation training modules. Procedural simulation training was first made commercially available through the RoSS simulator[218]. Yet to date validation of this simulator is limited to a single study showing the procedural training offered by the RoSS to be superior to console training[48]. This study provides evidence to support the utility of procedural VR simulation training and was the first comparison of basic and procedural robotic skills training. Tubes 2 module from Mimic (Mimic Technologies, Seattle, WA, USA) does also offer procedure specific training namely for UVA however it is more akin to a synthetic dry lab training model than full procedural training offered by the RobotiX mentor. This training has also only undergone limited validation[123]. In this chapter, evidence for construct validity for the procedural training modules of the RobotiX
mentor has been demonstrated across a number of domains. The study has shown procedural training with the RobotiX mentor to be an effective training tool for novice surgeons with a number of useful metrics for skill assessment. It should be noted that a number of metrics such as movements of each arm were found not to be reliable for training and could not easily differentiate novice, intermediate and expert participants. Overall the evidence does support the use of the simulator for training. On the other hand, the current metrics have not been shown to provide a reliable enough assessment of technical proficiency to be used in formal, high stakes skills testing. The limitations of these metrics should also be considered when evaluating trainee progression and outcomes.

The second aim of this study was to compare the effects of basic, procedural and no training on performance on a cadaveric skill transfer task. Further comparison of basic and procedural training was performed on a complex surgical task, a UVA on a dry lab model. The study benefitted from a large cohort of novice surgeons from 5 institutions with the trial conducted across 2 simulation training centres. Participant in either simulation underwent a programme of VR training distributed over a number of weeks. The results showed procedural training to be superior both to no training and basic VR training when assessed on cadavers. No significant difference in GEARS score was seen between procedural and basic VR training on the more complex UVA suture task although scores following procedural training were higher. These results are supported by systematic reviews of laparoscopic training showing procedural training to result in greater training outcomes[130, 9]. Although the cause for these differences in basic and procedural simulation for novice surgeons could not be ascertained, it can be hypothesised that procedural training triggers more interest and motivation even in novice surgeons resulting in greater concentration and improved learning outcomes. Whether such training is superior to focussed, deliberate practice needs to be investigated.
10.1.5 Comparative Analysis of Dry Lab & VR Simulation Training

Due to a number of factors largely unique to robotic surgery, basic technical skills training is predominantly undertaken using VR simulators. High fixed and variable costs of the Da Vinci system mean both that there are relatively few robotic systems, often limited to one per hospital, and the systems are generally used intensively to maximise their efficiency. Availability of robotic systems for dry lab training is therefore limited. VR simulators, whilst still expensive, and with the exception of the intuitive VR “backpack”, allow unrestricted training. A second benefit of VR simulation as demonstrated in chapter 3, p. 48, is the guided training and automatic, objective assessment that is thought to reduce the need for proctoring. Yet even with VR simulation, the majority of current training programmes continue to use surgical proctors rather than inbuilt VR software (see Chapter 2, p. 25). In comparison dry lab training requires more input from trainers.

As seen in Chapter 2, p. 25, the majority of robotic surgical training programmes use VR systems for basic skills training. Yet the results of this study have shown that whilst both modalities offer effective robotic technical skills, dry lab training does give specific benefits over VR simulators. No prior comparisons between dry and VR have been undertaken in robotic surgery but as discussed in the study, similar differences have been shown for laparoscopy. This study has important implications for robotic training. The study does support the use of either modality for basic robotic skills training. Yet where possible, dry lab training does offer useful advantages over VR supporting its inclusion in robotic curricula. In addition to higher marginal technical skills scores, dry lab training also scored higher in participant satisfaction outcomes. This highlights another important factor in simulation training. Trainee motivation is critically important
in ensuring successful training outcomes. A number of studies have similarly shown greater satisfaction with console training (i.e. dry lab, cadaveric and porcine training) in comparison to virtual reality[133, 164, 223, 225]. Although no direct comparison was performed during the study comparing procedural and basic simulation training, one of the main criticisms from the participants of the procedural training modules was of the texture of the tissue and the “floating sutures”. As hypothesised in this study, the inability of simulators to fully model the interactive physical environment results in both the reduced technical training and lower satisfaction levels.

10.1.6 Human Cadaveric Robotic Simulation Training

Human cadaveric training offers the highest fidelity simulation training. Difficulties highlighted in previous chapters with simulation fidelity and realism are largely avoided with cadavers which offer a realistic replication of the surgical environment and its anatomical idiosyncrasies. It provides the most accurate platform for practicing surgical techniques but is limited by the lack of bleeding and other physiological responses to surgery. Furthermore, ethical and legal considerations further limit its uptake. Extensive facilities and staff are required for preparation and storage of cadavers and costs of providing cadaveric training are high.

This study focussed on the cadaveric simulation training for robotic surgery. Prior to this study, no robotic cadaveric training programmes were being run in the UK and therefore I developed an evidence training protocol. Implementing the findings from Chapter 2, p. 25, a mixed methods programme was designed with the assistance of expert robotic surgeons and educationalists. The course consisted of knowledge-based learning delivered through small group teaching, surgical observation with live and as-live surgery and hands on technical skills training. Simulation training incorporated VR and cadavers with a focus on
cadaveric procedural training. Proctoring by expert robotic surgeons was provided to pairs of participants and over the two-day training course all participants gained experience in pelvic (bladder, prostate and lymph node dissection) and renal robotic procedures. For quality control and validation, all cadaver training was recorded via the robotic laparoscopic camera. Further qualitative feedback was recorded from participants. Evidence has been reported to support the validity of robotic cadaveric training. The fidelity and realism of the cadavers was supported by the positive qualitative data from participants supporting the course content. Training was also shown to be effective with significant improvements in technical performance irrespective of the robotic experience of participants.

This study has demonstrated the feasibility and acceptability of cadaveric training. When compared with the existing literature there is little doubt that cadavers offer high quality and effective simulation training. The main limitations result from the considerable logistical and financial barriers to offering such training. Understandably there are not obvious solutions to overcoming them and looking to the future cadaveric training will continue to remain greatly limited by them. As a result, careful consideration needs to be given as to how to most effectively incorporate cadaveric training into a robotic training curriculum. In this study, participation was limited to senior surgical trainees and consultant surgeons. A number of cadaveric courses for basic skills have been reported[13, 223]. Outcomes were uniformly positive but in spite of positive feedback from participants, it could not be shown that cadaveric training offered any significant benefits over VR or dry lab training. This study did not compare basic and procedural training but interestingly similar levels of improvement were seen irrespective of the prior experience of candidates. Nonetheless, data both in this thesis (see Chapters 3(p. 48), 4 (p. 69), 5, p. 110) and in prior research (Chapter 2 p. 25) show that basic skills training can be effectively performed using these bench top methods. It should be concluded that wet lab training is best reserved for experienced trainees,
enabling practice of the procedural and non-technical skills that require the fidelity offered by cadavers.

10.1.7 Live Porcine Training for Robotic Surgery

To investigate the role of live porcine training for robotic surgery I undertook a qualitative analysis of participants who attended the Minimal Invasiv Udviklings Centre, Aalborg University Hospital, Aalborg, Denmark. Due to legal restrictions, live animal training cannot be undertaken in the UK. All participants who attended the course were contact per email and ask to complete an online survey. Data were collected on prior surgical experience, self-assessed robotic surgical competency before and after completing the training course, outcomes from the training course and the feasibility and acceptability of training. The free-standing course was aimed at senior trainees and consultant surgeons wishing to extend and develop their robotic training. Training was open to all surgical specialities however training was predominantly focussed on urological surgery.

28% of participants completed the survey. Results demonstrated that despite the aims of providing training to more experienced surgeons, the majority had little or no prior robotic console experience. This was reflected in low self-assessed competency prior to training. Most participants had undertaken prior training (74%). Of those who had undertaken training, most had performed VR training (49%) while 21% had prior experience of wet lab or modular operating room experience. 41% had no hands-on robotic training experience having either no training experience at all (26%) or only e-learning (15%). The training was positively scored by the majority of participants for training outcomes and acceptability. Self-assessed competency and confidence scores also showed significant improvements following training (<0.0001 in each case). Following the course, 48% of participants continued robotic training train or went on to perform robotic surgery.
Interestingly post course self-assessed technical ability was not correlated with pre course self-rated ability but it did correlate with robotic surgical experience.

This study has shows that despite low pre-course robotic experience, porcine training was both effective in meeting the expectations of participants and in improving self-assessed technical ability and confidence. Importantly, a large portion of participants continued to train in robotic surgery with 52% undertaking OR training. This is surprising given the low levels of prior experience but supports the effectiveness of live porcine training. The limitations of these results need to be acknowledged but nonetheless they do support the inclusion of live porcine training in robotic training programmes.

10.1.8 Non-Technical Skills Training for Robotic Surgery

As a result of the increased awareness and training, there is little doubt that NTS are of critical importance for safe and successful surgery. Failures in team working and communication behaviours are the known to be root cause in the majority of surgical errors. Surgical teams present unique NTS challenges and errors need to be considered in the context of these complex team interactions. Major procedures like robotic surgery demand a variety of professionals working together with overlapping but differing skill sets, complex anaesthetic and surgical inventions, use of complex technology. All are undertaken in an environment prone to frequent distractions and interruptions. As discussed in Chapter 7, p. 142, a variety of training and assessment tools have been developed for NTS training and assessment. Yet the unique characteristics of robotic surgery warrant the development of a specific NTS behavioural rating system. A protocol was developed on the basis both of observation of live robotic surgical procedures, discussions with expert robotic surgeons and input from experienced surgeons, NTS experts and educationalists. The ICARS behavioural rating system has
been extensively validated in simulated operating room environment. Evidence to support the content, response process, internal structure and the relationships with other variables of the overall construct has been reported. Feedback from assessors confirmed the acceptability and feasibility for this novel rating system for robotic surgeons.

These results support the implementation of ICARS for NTS training in robotic surgery. Further work is required to build on this initial evaluation of ICARS and establish the validity of the tool for assessing NTS during live surgery. The specific implementation of NTS in a robotic surgical curriculum also needs to be addressed. Current evidence demonstrates that NTS training is effective when delivered as specific focussed training\[61, 43\]. There is even evidence to suggest that when combined with technical skills training, outcomes appear worse\[10\].

10.1.9 Motor Imagery Training for Robotic Surgery

Motor imagery training has the potential to offer a highly useful adjunct to training. It offers the possibility of enabling training outside of the simulation laboratory and to continue to develop the specific neural motor pathways and connectivity. Evidence of the benefit of MI for surgical training has reported previously and it is more widely employed for sports training and neuro-rehabilitation physiotherapy\[255\]. From surgical and non-surgical research, various factors have been identified for effective MI training\[58\]. Building on this data, this study reports the first randomised controlled trial of MI for robotic surgery using an evidence based and reproducible method to augment technical skills training. The study did use only novice surgeons, however the large number of participants recruited supports the validity of the results. Using a structure, evidence based approach to MI training resulted in significant improvements in technical ability compared to control. Use of the PETTLEP model was critical to the successful
10.2 Recommendations on Technical Skills Simulation Training for Robotic Surgery

outcomes of the trial. Studies by Davison et al. [58], Arora et al. [15] support the importance of effective training. Interestingly no effects on NTS were seen. MI remains an experimental training tool however the high-quality evidence from this study supports its inclusion in the robotic surgical curriculum.

10.2 Recommendations on Technical Skills Simulation Training for Robotic Surgery

This thesis has sort to analyse both the essential components of a simulation curriculum for robotic surgery and the necessary educational models of training. The objective of this body of work was to inform the development of a structured training programme for robotic surgery. Recommendations have focussed on the principle modalities of robotic assisted surgical training; VR, dry lab models; wet lab (cadaveric and live-animal). In addition to these technical training tools, non-technical skill and cognitive training tools were also evaluated for potential inclusion. An the basis of the results from this thesis, I have made the following recommendations on the design and structure of a simulation training curriculum for robotic surgery. This training curriculum will provide the necessary training to support surgeons in developing from beginner or novice to advanced trainees.
10.2 Recommendations on Technical Skills Simulation Training for Robotic Surgery

Fig. 10.1 The MARS Project Robotic Surgical Simulation Curriculum
10.2 Recommendations on Technical Skills Simulation Training for Robotic Surgery

The proposed robotic simulation curriculum offering a graduated programme of training is shown in Figure 10.1, p. 219. This outlines training for novice, intermediate and expert surgeons. For novice surgeons training should commence with VR or Dry lab based basic skills simulation training. On the basis of the systematic review (Chapter 2, p. 25) and the study of dry lab and VR simulation reported in Chapter5 (p. 110) the two training modalities offer very similar outcomes and both should be considered effective in basic skills training. In both cases competency-based training should be undertaken as demonstrated in Chapter 3 (p. 48). On the basis of this study, novice participants need to demonstrate competency in core robotic skills by achieving the benchmark scores in the basic skills exercises (Pick and Place, Camera Targeting 1 and Peg Board 1). Once competency in the generic skills has been attained, trainees are able to focus on reaching the competency standards in more complex tasks (thread the rings and suture sponge). A similar approach should be undertaken for dry-lab training with equivalent tasks being used to provide goal directed training. As demonstrated in Chapter 3 and supported by the results from my systematic review (Chapter 2, p. 25), competency as opposed to time based training supports a more focussed and efficient approach to skills development. Yet to date, whilst such techniques have been used in the majority of training programmes, albeit to a lesser extent in robotic surgery, they have not been incorporated into a robotic surgical curriculum. The other important consideration is the value of dry lab training. Where possible dry lab training should be offered in addition to VR training once basic competencies have been achieved to attain the benefits of real life training. Results from Chapter 4 (p. 69) has shown the benefits of procedural based VR training for novice trainees supporting its role alongside VR and dry-lab simulation. Like for dry lab training, competency based training programmes need development to support this training.
Once competency in basic skills has been achieved, trainees should move onto more complex training. Intermediate and expert trainees continue to benefit from basic skills training as shown by data from Chapter 2, p. 25, and Chapter 7, p. 142. Therefore these VR and dry lab training should remain important components of the simulation curriculum. Contrariwise the results published in this work should that both human cadaveric (Chapter 6, p. 124) and live porcine (Chapter 7, p. 142) wet lab training may also be beneficial to intermediate level trainees. As a result wet lab training is recommend for both intermediate and advance level trainees.

Alongside these simulation modalities, the importance of non-technical skills training needs to be recognised. Results from development and validation of the ICARS rating tool (Chapter 8, p. 163) have shown that it may applicable to both novice and experienced robotic surgeons. As has been reported previously, NTS is a vital component for all forms of surgical training. NTS training should be provided through robotic surgical training to ensure the equal development of NTS and technical skills. Uniquely I have also been able to demonstrate the role of a novel simulation training adjunct. As shown by the results reported in Chapter 9, p. 185, MI is an effective tool for supporting technical skills training. On this basis, it is recommended for all trainees from novice to advanced level.

These recommendations are summarised in Figure 10.1, p.219. The aim of this curriculum is to develop basic and advanced robotic technical and non technical skills. Completion of the curriculum would enable to undertake further modular training within the OR. Within the context of appropriate and tailor supervision, modular training could be instigated in parallel to the completion of simulation training curriculum, although integration of simulation and OR based training was beyond the scope of this thesis. As demonstrated by the international, multi-centre approach that was necessary to complete this thesis, provision of a comprehensive simulation training curriculum will exceed the capabilities of
individual training centres. Robotic simulation training requires access to both an array of high cost equipment, facilities and expertise. Consequently training should be organised nationally and even supranationally to ensure trainees have access to all components of the training curriculum.

10.3 Limitations and Future Work

The work contributing to the MARS project needs to be considered in light of a number of limitations. Firstly, the recommendations of the MARS projected are based on a series of independent studies. As a result, cross-over effects from concurrent training methods will not be detected. To fully validate the MARS project, ongoing data collection needs to be undertaken to ensure that curriculum is delivery the training aims as expected. Aside from validation of the Porcine Training programme, data was collected within the UK with the majority of participants completing training within the UK medical training programme. Variations in training structures in other country may impact the recommendations of this body of work. Finally, the MARS project is limited to evaluating the various components of the simulation curriculum. Modular OR training offers trainees the opportunity to develop robotic skills safely within the clinical environment. Whilst it was outside the remit of the MAR project, the transition from simulation lab to the OR needs further evaluation. Finally the cost and time of implications of training need to be formally evaluated.
REFERENCES


References


Systematic Review Search Protocol

The following search strategy was used for medline query and an equivalent protocol was used for EMBASE.

((simulat*[Title/Abstract] AND surgery[Title/Abstract] AND curricul*[Title/Abstract])
OR (simulat*[Title/Abstract] AND surgery[Title/Abstract] AND program*[Title/Abstract])
OR (simulat*[Title/Abstract] AND surgery[Title/Abstract] AND training[Title/Abstract]
AND program*[Title/Abstract]) OR (surgery[Title/Abstract] AND training[Title/Abstract]
AND program*[Title/Abstract]) OR (surgery[Title/Abstract] AND training[Title/Abstract]
AND simulat*[Title/Abstract]))
Appendix B

A Comparative Analysis of Virtual Reality and Dry-Lab Robotic Simulation Training: Rater Training and Validation

In order to provide technical assessment using the validated GEARS global rating scale I undertook a period of dedicated technical assessment training. Building on my experience of assisting in over 100 robotic surgical procedures, initially I watched over 100 of video recordings of robotic surgery procedures in addition to simulation training performances. All videos had previously been assessed by expert surgeons using GEARS and I compared these assessments to recorded performances. In addition I carefully studied the development and validation of GEARS.

Following this period of training, I performed a number of test assessments. I blindly assessed 19 video recordings of surgeons of varying levels of experienced were marked using GEARS and then compared to assessments by an expert surgeon. A two way mixed model was used to test for agreement. This demonstrated an intraclass correlation coefficient (ICC) of 0.623, p<0.0005 indicating a substantial agreement[129]. I compared and contrasted my assessment to those of expert surgeons. Following this period of review and analysis, I repeated the process with
a further 9 videos of simulated robotic surgery performed by various surgeons of varying experience. This demonstrated a further improvement in my assessment in comparison to an expert surgeon with an ICC of 0.67. This results demonstrate my ability as an accurate and consistent GEARS rater and validate my ability to assess technical robotic skill.
APPENDIX C

CADAVERIC WET LAB SIMULATION TRAINING: SUPPLEMENTARY INFORMATION

C.1 Further Statistical Analysis

Normality of assessment of data was undertaken through using visual analysis of histograms (Figure C.1), skewness and kurtosis analysis and Shapiro-Wilk’s ($W = 0.95$, $p = 0.002$). On this basis data was determined to be non-normal.

![Fig. C.1 Histogram of Total GEARS Score](image-url)
C.2 Example Programme for the Cadaveric Simulation Training Course for Robotic Surgery

![Normal P-P Plot](image)

**Fig. C.2 Normal P-P Plot**

C.2 Example Programme for the Cadaveric Simulation Training Course for Robotic Surgery
### Estimates of Fixed Effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>14.54</td>
<td>.455</td>
<td>50.1</td>
<td>31.9</td>
<td>.000</td>
<td>13.62 - 15.456</td>
</tr>
<tr>
<td></td>
<td>2387</td>
<td>105</td>
<td>88</td>
<td>54</td>
<td></td>
<td>8366 - 408</td>
</tr>
<tr>
<td>[TimeNumeri</td>
<td>-2.85</td>
<td>.447</td>
<td>23.0</td>
<td>-6.3</td>
<td>.000</td>
<td>-3.779 - -1.929</td>
</tr>
<tr>
<td>c=1.00]</td>
<td>4085</td>
<td>154</td>
<td>36</td>
<td>83</td>
<td></td>
<td>014 - 156</td>
</tr>
<tr>
<td>[TimeNumeri</td>
<td>0^b</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>c=2.00]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[TimeNumeri</td>
<td>.1119</td>
<td>.028</td>
<td>43</td>
<td>3.99</td>
<td>.000</td>
<td>.0554 - .16839</td>
</tr>
<tr>
<td>c=1.00]  *</td>
<td>43</td>
<td>000</td>
<td>49</td>
<td>8</td>
<td></td>
<td>93 - 2</td>
</tr>
<tr>
<td>RobotictExperiencesCases</td>
<td>.0625</td>
<td>.029</td>
<td>48.6</td>
<td>2.11</td>
<td>.040</td>
<td>.0030 - .12202</td>
</tr>
<tr>
<td>c=2.00]  *</td>
<td>38</td>
<td>595</td>
<td>92</td>
<td>3</td>
<td></td>
<td>54 - 2</td>
</tr>
<tr>
<td>[TimeNumeri</td>
<td>.2978</td>
<td>.108</td>
<td>43.5</td>
<td>2.75</td>
<td>.008</td>
<td>.0800 - .51556</td>
</tr>
<tr>
<td>c=1.00]  *</td>
<td>06</td>
<td>019</td>
<td>72</td>
<td>7</td>
<td></td>
<td>47 - 5</td>
</tr>
<tr>
<td>DryLab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[TimeNumeri</td>
<td>.2494</td>
<td>.142</td>
<td>57.9</td>
<td>1.75</td>
<td>.084</td>
<td>-.0349 - .53388</td>
</tr>
<tr>
<td>c=2.00]  *</td>
<td>90</td>
<td>074</td>
<td>88</td>
<td>6</td>
<td></td>
<td>04 - 4</td>
</tr>
</tbody>
</table>

Fig. C.3 Linear mixed model assessing pre-post GEARS score controlling for dry lab simulation and robotic surgical experience
ROBOT-ASSISTED UROLOGICAL SURGERY TRAINING

CADAVERIC & SIMULATION TRAINING COURSE

July 31st & August 1st 2015

12 CPD Points Available

Guys Hospital, London
Kings College London

Course Directors: Prof. Prokar Dasgupta
Mr M Shamim Khan

Course Coordinators: Mr Kamran Ahmed
Mr Nicholas Raison

Fig. C.4 Example Programme for the Cadaveric Simulation Training Course for Robotic Surgery
## Faculty

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark Speakman</td>
<td>Consultant Urologist, President, BAUS</td>
</tr>
<tr>
<td>Christian Brown</td>
<td>Consultant Urologist, King’s College Hospital &amp; Guy’s Hospital</td>
</tr>
<tr>
<td>Ben Challacombe</td>
<td>Consultant Urologist, Guy’s Hospital</td>
</tr>
<tr>
<td>Khurshid Guru</td>
<td>Roswell Park Cancer Institute, NY, US</td>
</tr>
<tr>
<td>Senthil Nathan</td>
<td>Consultant Urologist, UCLH</td>
</tr>
<tr>
<td>Naeem Soomro</td>
<td>Consultant Urologist, Newcastle upon Tyne Hospitals NHS Trust</td>
</tr>
<tr>
<td>Ramesh Thurairaja</td>
<td>Consultant Urologist, Guy’s Hospital</td>
</tr>
<tr>
<td>Prokar Dasgupta</td>
<td>Professor, Consultant Urologist</td>
</tr>
<tr>
<td>M Shamim Khan</td>
<td>Reader, Consultant Urologist</td>
</tr>
<tr>
<td>Kamran Ahmed</td>
<td>NIHR Academic Clinical Lecturer / SPR in Urology</td>
</tr>
<tr>
<td>Nicholas Raison</td>
<td>Vattikuti Research Fellow</td>
</tr>
<tr>
<td>Maria Nightingale</td>
<td>Urology Operating Theatre Sister</td>
</tr>
</tbody>
</table>

Fig. C.5 Example Programme (cont’d)
### Lecture Programme

**Friday 31\textsuperscript{st} July**

<table>
<thead>
<tr>
<th>Time</th>
<th>Friday 31\textsuperscript{st} July</th>
<th>Faculty</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>0830</td>
<td>Registration and Introduction</td>
<td></td>
<td>Gordon Museum</td>
</tr>
<tr>
<td>0900</td>
<td><strong>Basics of Robotics</strong></td>
<td><strong>Intuitive</strong></td>
<td>Gordon Museum</td>
</tr>
<tr>
<td></td>
<td>Introduction to the robot, Docking and Trouble</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0930</td>
<td><strong>Robotic Surgery in the NHS</strong></td>
<td>Mr Speakman</td>
<td>Gordon Museum</td>
</tr>
<tr>
<td>1000</td>
<td>Port Placement</td>
<td></td>
<td>Gordon Museum</td>
</tr>
<tr>
<td>1030</td>
<td><strong>BREAK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>Robotic Prostatectomy- tips and tricks</td>
<td>Mr Brown</td>
<td>Gordon Museum</td>
</tr>
<tr>
<td>1130</td>
<td>Partial Nephrectomy</td>
<td>Mr Soomro</td>
<td>Gordon Museum</td>
</tr>
<tr>
<td>1200</td>
<td>Nephroureterectomy</td>
<td>Mr Nathan</td>
<td>Gordon Museum</td>
</tr>
<tr>
<td>1230</td>
<td><strong>LUNCH</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1330</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1400</td>
<td><strong>LIVE SURGERY: Robotic Assisted Radical Prostatectomy</strong></td>
<td>Mr Challacombe</td>
<td>Lecture Theatre</td>
</tr>
<tr>
<td>1430</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500</td>
<td><strong>BREAK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1530</td>
<td><strong>LIVE SURGERY: Robotic Assisted Radical Prostatectomy</strong></td>
<td>Mr Brown</td>
<td>Lecture Theatre</td>
</tr>
<tr>
<td>1600</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1630</td>
<td>Final Questions</td>
<td>All Faculty</td>
<td>Gordon Museum</td>
</tr>
</tbody>
</table>

Fig. C.6 Example Programme (cont’d)
### Lecture Programme

**Saturday 1st August**

<table>
<thead>
<tr>
<th>Time</th>
<th>Saturday 1st August</th>
<th>Faculty</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>0830</td>
<td>Introduction</td>
<td>TBC</td>
<td>Gordon Museum</td>
</tr>
<tr>
<td>0900</td>
<td>Cystectomy</td>
<td>Mr Thurairaja</td>
<td>Gordon Museum</td>
</tr>
<tr>
<td>0930</td>
<td><strong>LIVE SURGERY: Robotic Assisted Radical Cystectomy</strong></td>
<td>Mr Khan</td>
<td>Lecture Theatre</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1030</td>
<td><strong>BREAK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1130</td>
<td><strong>LIVE SURGERY: Robotic Assisted Radical Cystectomy</strong></td>
<td>Mr Khan</td>
<td>Lecture Theatre</td>
</tr>
<tr>
<td>1200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1230</td>
<td><strong>LUNCH</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1330</td>
<td>Tips and Tricks in Cystectomy</td>
<td>Mr Guru</td>
<td>Lecture Theatre</td>
</tr>
<tr>
<td>1400</td>
<td><strong>LIVE SURGERY: Robotic Assisted Radical Cystectomy</strong></td>
<td>Mr</td>
<td>Lecture Theatre</td>
</tr>
<tr>
<td>1430</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500</td>
<td><strong>BREAK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1530</td>
<td><strong>LIVE SURGERY: Robotic Assisted Radical Cystectomy</strong></td>
<td>Mr B Challacombe</td>
<td>Lecture Theatre</td>
</tr>
<tr>
<td>1600</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1630</td>
<td><strong>Final Questions</strong></td>
<td>All Faculty</td>
<td>Lecture Theatre</td>
</tr>
</tbody>
</table>

Fig. C.7 Example Programme (cont’d)
### Robotic Virtual Reality Simulation
**Friday 31st July**

<table>
<thead>
<tr>
<th>Time</th>
<th>Simulator 1</th>
<th>Simulator 2</th>
<th>Venue</th>
</tr>
</thead>
<tbody>
<tr>
<td>0900-0930</td>
<td>Delegate 1</td>
<td>Delegate 2</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td>0930-1000</td>
<td>Delegate 3</td>
<td>Delegate 4</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td>1000-1030</td>
<td>Delegate 5</td>
<td>Delegate 6</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td>1030-1100</td>
<td>Delegate 7</td>
<td>Delegate 8</td>
<td>Lower Dissecting Room</td>
</tr>
</tbody>
</table>

### Robotic Cadaveric Simulation
**Friday 31st July**

<table>
<thead>
<tr>
<th>Time</th>
<th>DaVinci Xi</th>
<th>Venue</th>
</tr>
</thead>
<tbody>
<tr>
<td>0930-1100</td>
<td>Delegate 1</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 2</td>
<td></td>
</tr>
<tr>
<td>1100-1230</td>
<td>Delegate 3</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 4</td>
<td></td>
</tr>
<tr>
<td>1230-1400</td>
<td>Delegate 5</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 6</td>
<td></td>
</tr>
<tr>
<td>1400-1530</td>
<td>Delegate 7</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 8</td>
<td></td>
</tr>
<tr>
<td>1530-1700</td>
<td>Delegate 9</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 10</td>
<td></td>
</tr>
</tbody>
</table>

Fig. C.8 Example Programme (cont’d)
### Robotic Cadaveric Simulation
Saturday 1st August

<table>
<thead>
<tr>
<th>Time</th>
<th>DaVinci Xi</th>
<th>Venue</th>
</tr>
</thead>
<tbody>
<tr>
<td>0930-1030</td>
<td>Delegate 1</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 2</td>
<td></td>
</tr>
<tr>
<td>1030-1130</td>
<td>Delegate 3</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 4</td>
<td></td>
</tr>
<tr>
<td>1130-1230</td>
<td>Delegate 5</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 6</td>
<td></td>
</tr>
<tr>
<td>1230-1330</td>
<td>Delegate 7</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 8</td>
<td></td>
</tr>
<tr>
<td>1330-1430</td>
<td>Delegate 9</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 10</td>
<td></td>
</tr>
<tr>
<td>1430-1530</td>
<td>Delegate 12</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td></td>
<td>Delegate 13</td>
<td></td>
</tr>
</tbody>
</table>

### Robotic Cadaveric Simulation
Saturday 1st August

<table>
<thead>
<tr>
<th>Time</th>
<th>Simulator 1</th>
<th>Simulator 2</th>
<th>Venue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1445-1515</td>
<td>Delegate 9</td>
<td>Delegate 10</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td>1515-1545</td>
<td>Delegate 11</td>
<td>Delegate 12</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td>1545-1615</td>
<td>Delegate 5</td>
<td>Delegate 6</td>
<td>Lower Dissecting Room</td>
</tr>
<tr>
<td>1616-1645</td>
<td>Delegate 7</td>
<td>Delegate 8</td>
<td>Lower Dissecting Room</td>
</tr>
</tbody>
</table>

Fig. C.9 Example Programme (cont’d)
Venue and Arrangements

Venue

Anatomy Laboratory and Gordon Museum
Hodgkin Building
King’s College London
Guy’s Campus
London
SE1 1UL

Travel

Guy’s campus is located in central London adjacent to London Bridge Station (national rail and underground station).

• Driving to Guy’s
There are a number of private car parks within the vicinity however it should also be noted that the campus is within the congestion charge zone. If you are planning to drive, please contact us for help with organising parking.

• Travelling by Train
Guy’s campus is adjacent to London Bridge Station (national rail and London underground links).

• Flying to London
London is served by 4 main airports:
  ○ • Heathrow with underground connection and direct trains to London Paddington
  ○ • Gatwick with direct train connections to London Bridge, Victoria and St Pancras
  ○ • Luton with direct train links to St Pancras and Blackfriars
  ○ • Stansted with direct train links to Liverpool Street

Course Dinner

The faculty dinner will be held on
Friday 31st August.
(Details to be confirmed)
Appendix D

Live Porcine Robotic Training Course Participant Survey Questionnaire
1. Level of training at the time of completing the Aalborg course (most recent if multiple visits)

2. Your Speciality

3. In which country do you practice?

4. Before completing the course, how would you have rated your robotic surgical skill?
   - Complete novice
   - Expert robotic surgeon

5. Before completing the course, how confident were you in performing robotic surgery?
   - Not at all confident
   - Completely confident

6. Before the course, approximately how many robotic procedures had you been involved in as console surgeon?
   - 0
   - 100+

7. Before the course, had you undertaken any other training in robotic surgery?
   - e-Learning
   - Virtual Reality Simulation
   - Robotic training using plastic models
   - Cadaveric training
   - Live animal training
   - Modular operating room training

---

Fig. D.1 Aalborg Participant Survey Questionnaire
8. Following the course, how would you rate your surgical ability?

- Complete novice
- Expert robotic surgeon

9. Following the training course, how confident did you feel in your robotic surgical skill

- Not at all confident
- Completely confident

10. Overall how would you rate the educational benefit of the course?

- Not beneficial
- Extremely beneficial

11. How effective do you think the live porcine robotic course was in training robotic skill?

- Not at all effective
- Extremely effective

12. How useful were the different modules of the course?

<table>
<thead>
<tr>
<th>Not Performed</th>
<th>Not useful</th>
<th>Slightly useful</th>
<th>Moderately useful</th>
<th>Very useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port placement and robotic docking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic robotic skills (endowrist manipulation, camera control etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repair of bladder injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pelvis Lymph Node Dissection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ureteric re-implantation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nephrectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial Nephrectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyeloplasty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salpingectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hysterectomy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. D.2 Aalborg Participant Survey Questionnaire (cont’d)
Fig. D.3 Aalborg Participant Survey Questionnaire (cont’d)
APPENDIX E

ICARS DEVELOPMENT SUPPLEMENTARY DATA

Fig. E.1 Comparison of Novice, Intermediate and Expert ICARS Performance Scores